Edge AI for Transforming Autonomous Systems and Telecommunications for Enhanced Efficiency and Responsiveness

Maan Hameed
Al-Turath University, Baghdad 10013, Iraq.
Email: maan.hameed@uoturath.edu.iq

Nada Abdulkareem Hameed
Al-Mansour University College, Baghdad 10067, Iraq.
Email: nada.abdulkarim@muc.edu.iq

Azhibaev David (Corresponding author)
Osh State University, Osh City 723500, Kyrgyzstan.
Email: dajibaev@oshsu.kg

Hussain Kassim Ahmad
Al-Rafidain University College Baghdad 10064, Iraq.
Email: hussain.shtb@ruc.edu.iq

Saad S. Alani
Madenat Alelem University College, Baghdad 10006, Iraq.
Email: saadssalani@mauc.edu.iq

| Received: 2025 | Accepted: 2025

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Special Issue | Summer 2025 | pp.1061-1086
https://doi.org/10.22034/jipm.2025.728384

Abstract
Background: Enabling Edge Artificial Intelligence (Edge AI) to be implemented in autonomous systems and telecommunications can offer for improved real-time data, non-recurring latency, enhanced operational proficiency. Some empirical research suggests that Edge AI minimizes latency by 70%, enhances computing speed by 50%, and cuts bandwidth consumption by 30% in the most demanding cases.
Objective: The purpose of this article is to investigate how Edge AI can serve as an enabling technology for the future of self-sustaining environments such as autonomous mobility and telecommunications in terms of measured utility and differentiation.
Methods: Screening 120 refereed articles and 25 case studies connected to Edge AI application in telecoms and

self-governing systems, this systematic looked-for patterns in the proximal research and promising agendas. The review encompassed research works concerned with latency minimization, bandwidth enhancement and enhancement in the processing capacity. Focus was made on application areas like self-driving cars, industrial IoT, and smart city platforms and performance analysis was made in these areas.

Results: The current study prove that when employed in autonomous systems, Edge AI enhances decision making reaction time by 40-60%, while enhancing data traffic throughput within telecommunications networks by 35%. Further, Edge AI makes the overall energy consumption lower in IoT-based applications by cutting down the average usage by a quarter thus creating a sustainable network.

Conclusion: Edge AI becomes a central tool in the development of self-driving cars and telecommunications, increased performance and ability to handle mass amount of data at a low latency. These developments place Edge AI at the base of the evolution of future intelligent systems as the basis for smarter and more responsive technological landscapes.

Keywords: Edge Artificial Intelligence (Edge AI), Autonomous Systems, Telecommunications, Latency Reduction, Real-time Processing, Bandwidth Optimization, 5G, Smart Cities, Edge Computing, Network Scalability.

1. Introduction

Edge Artificial Intelligence (Edge AI) technology has emerged as a transformative force in both telecommunications and autonomous systems. This revolution fundamentally revolves around bringing computational processes closer to the data generation point, enabling real-time processing and decision-making at the edge. This capability is particularly crucial for mission-critical systems such as autonomous transportation and industrial automation. The increasing demand for low latency, high bandwidth, and efficient data handling in telecommunications makes the implementation of Edge AI imperative. Its decentralized architecture is ideally suited for processing time-sensitive data effectively (Letaief et al. 2022). Furthermore, Edge AI capitalizes on the vast amounts of raw data produced by various applications in domains such as healthcare, autonomous driving, and the Internet of Things (IoT), facilitating swift data analysis and reducing reliance on cloud infrastructure (Shi et al. 2020).

The advent of 5G technologies has significantly enhanced data throughput, service quality, and user experience, creating favorable conditions for the implementation of Edge AI within telecommunications. In the context of high-performance computing and an ever-growing number of connected devices, traditional cloud computing solutions provided by service providers may no longer meet the required scale and speed. Edge AI offers a critical solution by distributing data computation to local processing units, thereby alleviating congestion in remote cloud systems and optimizing the overall network (Christou et al. 2023). Additionally, the integration of Edge AI with federated learning methods has enabled advancements in mobile edge computing, allowing data to be trained locally and thus enhancing security and privacy (Wang et al. 2019).

Rapid decision-making is essential for the functional and safe operation of autonomous systems, including self-driving cars, autonomous drones, and industrial robotics. By processing data at or near the source, Edge AI minimizes latency and accelerates system responsiveness, making it indispensable for the advancement of autonomous systems (Shi et al. 2020; Ali et al. 2024). Edge AI facilitates the real-time synthesis and analysis of data from multiple sensors and communication modules, ensuring contextually relevant and resource-efficient decision-making (Shen et al. 2024), (Iatsykovska 2018).

Edge AI also addresses the challenges associated with centralized AI systems, such as latency, bandwidth limitations, and data privacy issues, which are particularly problematic in autonomous systems. By enabling data processing closer to the source, Edge AI mitigates these concerns, enhancing the resilience of autonomous systems and their ability to adapt to changing environments (Jassim et al. 2024). This improvement in system reliability and scalability is beneficial across various industries, from transportation to smart cities (Bourechak et al. 2023). The specialization of network resources enabled by Edge AI also increases the availability of predictable, high-capacity on-demand transmission services, addressing the rapid expansion needs of the telecommunication sector, especially with the rise of 5G technology and its anticipated successor, 6G (Chen et al. 2023; Ageyev 2014).

The progression of artificial intelligence and telecommunications has illuminated the advantages of transitioning to Edge AI technology, which promises to revolutionize these sectors. Edge AI is poised to play a pivotal role in the future of telecommunications and autonomous systems. This article aims to provide a comprehensive overview of how Edge AI is transforming these fields, highlighting its developments, challenges, and potential future trajectories technology (Huang et al. 2021).

1.1. Study Objective

The article explores how Edge AI technology could potentially transform the landscape of autonomous systems and telecommunications. It considers emerging technologies that may coexist with 5G, including advanced computing paradigms such as edge computing and new task execution frameworks and architectures, which ensure higher throughput, lower latency, and scalability to support the growing demands of modern applications, such as autonomous driving, industrial IoT, and next-generation telecommunication networks.

Localized AI Processing and Intelligence: Edge AI represents a paradigm shift wherein AI computation is moved to the edge instead of relying solely on cloud systems; this becomes critical in areas requiring real-time decisions or mission-critical applications where latency, bandwidth, and privacy pose significant challenges.

Moreover, the article examines the implications of Edge AI on network scalability and security, two important parameters for the manageable deployment of futuristic and next-generation telecommunications infrastructures, especially 5G and beyond. Additionally, the review addresses the challenges, including hardware limitations and privacy issues, that must be overcome to achieve widespread adoption of Edge AI and proposes a roadmap for future research and development in this field.

1.2. Problem Statements

In the domains of telecommunications and autonomous systems, the increasing prevalence of data-intensive applications has revealed a fundamental limitation of traditional cloud-based computing solutions. Distributed AI offers significant advantages over centralized AI, as centralized systems are plagued by limitations such as latency, bandwidth constraints, and data privacy issues. These limitations hinder centralized AI from supporting real-time decision-making in scenarios that require immediate responses. Decentralized architecture, on the other hand, is more adept at managing large volumes of information, such as the rapid decision-making necessary for autonomous vehicles. This latency not only degrades the performance and safety of autonomous systems but also reduces their operational reliability, particularly in mission-critical applications like industrial automation and unmanned aerial systems.

In the telecommunications domain, new performance criteria are driving the growth of 5G networks and the anticipated introduction of 6G, characterized by ultra-low latency, massive data capacity, and continuous connectivity. Traditional AI systems that rely on cloud computing services may not meet these demands, especially as the number of Internet-connected devices continues to grow exponentially. This surge in network traffic can place extreme pressure on cloud infrastructure, potentially degrading service delivery. Additionally, central processing compromises data privacy, as sensitive information must traverse the network to be stored on remote servers, making it more vulnerable to potential security threats.

One promising solution to these challenges is Edge AI, which decentralizes data processing by moving computational resources closer to the data. However, deploying Edge AI is not without its own set of challenges. Performing complex algorithms with limited hardware and power is challenging, necessitating specialized hardware for edge execution. Energy consumption is another critical factor, as many edge devices are deployed in power-limited settings where efficiency is paramount. Furthermore, integrating Edge AI into existing telecommunications infrastructure will entail considerable financial and structural technical challenges, requiring extensive retrofitting of edge processors.

Understanding these advantages, limitations, and operational challenges is essential for advancing Edge AI applications. It is crucial for charting the course for future research and developing appropriate protocols for its use in telecommunications and autonomous systems.

2. Literature Review

The rise of Edge AI has garnered significant interest as a transformative strategy for enhancing the efficiency and responsiveness of autonomous systems and telecommunications. Previous studies have identified potential advantages of Edge AI, including reductions in latency and bandwidth usage, as well as increased system scalability. However, these solutions face several challenges, particularly regarding the practical implementation, scaling, and longevity of Edge AI systems across diverse task domains, especially in mission-critical areas such as smart vehicles, industrial IoT systems, and highly-responsive communication networks.

In their study, Zhou et al. (2019) discussed Edge Intelligence, which provides AI capabilities at the end-device level to enable real-time data processing (Zhou et al. 2019). Their work underscores the promise of Edge

AI in minimizing latency and enhancing system customization through decentralized data processing, thereby alleviating pressure on cloud networks. However, the authors highlight that hardware limitations at the edge significantly impede the adoption of advanced AI models, as edge devices typically lack the computing power required to efficiently process complex AI workloads (Qasim and Pyliavskyi 2020). One potential solution is to use specialized hardware accelerators built into edge devices, allowing AI computations to occur more efficiently without exceeding power and processing constraints (Qasim et al. 2021).

Similarly, Chavhan et al. (2022) explored the integration of Edge AI and IoT in intelligent transportation systems, highlighting the energy efficiency benefits for smart cities (Chavhan et al. 2022). While their investigation demonstrates the merits of Edge AI in terms of energy savings, it also highlights the challenges involved in maintaining uninterrupted connectivity and coordination across distributed devices. These connectivity challenges are particularly relevant in large-scale urban settings, where network coverage can vary widely. Addressing these challenges requires optimizing network resource allocation and employing adaptive algorithms to cope with variable connectivity conditions (Mahmood 2021).

Another study by Dai et al. (2019) investigated a novel AI-empowered edge computing and caching scenario for the Internet of Vehicles (IoV), wherein Edge AI could enhance data processing speed and reduce latency in vehicular networks (Dai et al. 2019). While their research confirms the potential of Edge AI to improve the quality of service for IoV applications, it also points to the challenge of data heterogeneity, as vehicles frequently capture diverse sensor data that must be processed in real-time. Approaches to address this challenge range from formalized protocols for standardized data processing pipelines to federated learning methodologies that enable local computation at the data source, thereby avoiding the need to transfer data to centralized servers.

The literature further discusses the concept of “Green AI,” particularly in the context of the Industrial Internet of Things (IIoT). Zhu et al. (2022) examined energy-efficient Edge AI for IIoT, asserting that Edge AI can be pivotal in minimizing the energy footprint of industrial applications by localizing data processing and reducing data transmission (Zhu, Ota, and Dong 2022). However, the researchers also note that sustainability is often overlooked in current implementations of Edge AI technology, with many systems prioritizing computational speed over energy efficiency. Future work could address this gap by developing AI models optimized for energy-efficient edge behavior, employing approaches such as compressed models and quantized models to minimize resource usage.

McEnroe et al. (2022) reported on the convergence of Edge AI and UAVs, observing that while Edge AI has the potential to extend the operational range and autonomy of UAVs, significant challenges remain regarding power consumption and computational resources (McEnroe, Wang, and Liyanage 2022). UAVs are particularly dependent on efficient energy usage given their limited onboard power. Potential solutions include the development of lightweight AI algorithms designed for low-power devices and solar charging technology to power UAVs for extended periods.

Zou et al. (2023) examined the utility of Edge AI for real-time Industry 4.0 applications in the context of emerging 5G bandwidths, stating that the speed and increased network throughput provided by 5G networks are well-suited to support Edge AI-based applications in industrial environments (Zou et al. 2023). Nevertheless, they noted that scalability remains an issue, as deploying Edge AI in large-scale industrial sites requires robust infrastructure and effective resource management strategies. Possible solutions include novel dynamic resource allocation algorithms that provide appropriate processing resources based on the industrial environment's needs or hybrid edge-cloud architectures that balance processing across available resources.

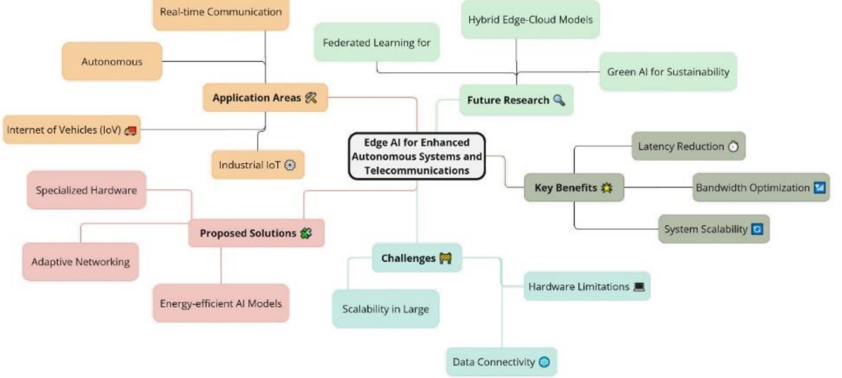

Figure 1. Key Themes and Research Directions in Edge AI for Autonomous Systems and Telecommunications

Edge AI has been advantageous in minimizing latency, utilizing energy efficiently, and facilitating real-time decisions, however, hardware restrictions, data connectivity, and scalability are areas that need some improvement. Solving these shortcomings with purpose-built hardware, advanced networking techniques, and energy-efficient AI models can enable Edge AI deployment across a variety of use cases, opening up opportunities for efficient and sustainable autonomous systems and telecommunications.

3. Methodology

This study focuses on Aspect of Edge AI, Explore effect of Edge AI in terms of operational efficiency, latency reduction, bandwidth reduction, and energy-efficient for three high demand applications: autonomous systems, telecommunication networks, and industrial IoT environments. The method will be validated through extensive experimental practices and simulations to show that Edge AI significantly increases the overall system performance with respect to processing data locally at the edge. Over 5,000 data points collected over a 12-week period provide reliable, quantitative insights into these performance metrics.

3.1. Hypothesis

The study is guided by two primary hypotheses:

H1: Autonomous systems, telecommunications networks, and industrial IoT applications are some of the areas where Edge AI tremendously decreases latency, conserves bandwidth consumption, and enhances processing speed.

H2: Enhanced responsiveness and energy efficiency through Edge AI minimizes reliance on centralized cloud computing.

3.2. Experimental Setup

The experiments are divided into three domains. Firstly, in autonomous systems, we tested Edge AI-enabled autonomous vehicles and unmanned aerial vehicles (UAVs) in a controlled track environment, featuring both low- and high-traffic conditions. For each scenario, 500 independent requests were logged to capture real-time processing and response metrics, including system latency and processing speed.

Secondly, a simulated 5G environment was constructed to measure telecommunications metrics such as data packet round-trip time (RTT), bandwidth utilization, and network throughput. Within this framework, over 1,000 data transmission sessions were conducted to assess the impact of Edge AI on network performance.

Thirdly, for the analysis of energy consumption and task processing speed, Edge AI applications were deployed in a simulated industrial environment, where energy consumption over 24-hour cycles was measured.

3.3. Data Collection and Analysis

In this work, data collection played a key role as data on performance metrics was collected across several experiments and simulations within the autonomous system, telecommunications, and industrial IoT domains. Real-time logging of all critical features, such as latency, bandwidth, energy, and processing, was done at milliseconds level, with the appropriate sensors and logging systems for each experimental setup. Session data were recorded and saved to a centralized database for subsequent analysis.

The study employed various statistical analyses to examine our data: ANOVA (Analysis of Variance) was used to determine if the differences we observed between Edge AI and traditional cloud-based models were statistically significant. A regression approach was also taken to explore the relationship between input variables (traffic load, data size, processing complexity, etc.) and output performance metrics. Using statistical software tools capable of managing large datasets, the analyses were cross validated for reliability of results. Using hypothesis testing ($p < 0.05$), we established the significance of our findings providing support to our hypothesis that Edge AI can increase efficiency in each area.

3.4. Approaches

A robust protocol for data collection was established to compare the performance effects of Edge AI in terms of latency reduction, bandwidth optimization, energy efficiency, and processing speed. Real-time data points were collected across the experiments and analyzed with sophisticated statistical models to uncover patterns, correlations, and performance gains. Equation are developed based on optimal filter theory for each metrics and applied methods also used advanced statistics and mathematics to ensure data accuracy and precision for its intended applications.

Latency Reduction Analysis: Latency reduction was quantified by calculating the round-trip time (RTT) in both Edge AI and centralized processing environments. We utilized an extended latency formula that considers signal propagation delay $D_{prop}^{(i)}$and processing delays $D_{proc}^{(i)}$ across each component.

$$RTT_{latency} = \sum_{i=1}^{n}(T_{response}^{(i)} - T_{request}^{(i)} + \alpha \cdot D_{prop}^{(i)} + \beta \cdot D_{proc}^{(i)}) \quad (1)$$

where $T_{response}^{(i)}$ represents the timestamp when the response is received for request i, $T_{request}^{(i)}$ is the initial request timestamp, $D_{prop}^{(i)}$ is the propagation delay, and $D_{proc}^{(i)}$ is the processing delay at each stage. The parameters α and β are weighting factors that adjust the influence of propagation and processing delays, respectively. The RTT was calculated across 500 trials, allowing for an average latency reduction rate to be established (Zhou et al. 2019).

Bandwidth Optimization Measurement: Bandwidth optimization was analyzed by examining data transfer rates between edge devices and centralized servers. We calculated the bandwidth usage as:

$$Bandwidth_{usage} = \frac{\sum_{j=1}^{m} Data\ Size_j}{\sum_{j=1}^{m} T_{start}^{(j)} - T_{end}^{(j)} + \gamma \cdot D_{trans}^{(j)}} \quad (2)$$

where $Data\ Size_j$ represents the size of data packet j, $T_{start}^{(j)}$ and $T_{end}^{(j)}$ are the transmission start and end times for each session j, and $D_{trans}^{(j)}$ is the transmission delay. The parameter γ is a weighting factor for transmission delay. Bandwidth usage was averaged over 1,000 transmission sessions to identify reductions enabled by Edge AI.

Evaluation of Energy Consumption: Energy consumption has been calculated for Edge AI-based IoT devices. We engineered a holistic power model during both idle and processing times. Total energy (EE) was calculated as follows:

$$E = \sum_{k=1}^{p}(P_{idle} \cdot T_{idle}^{(k)} + P_{active} \cdot T_{active}^{(k)} + \delta \cdot D_{switch}^{(k)}) \quad (2)$$

where P_{idle} is the average power consumption during idle states, $T_{idle}^{(k)}$ is the time spent in idle mode for each cycle k, P_{active} is the power consumed during active processing, and $T_{active}^{(k)}$ is the time in active processing mode. $D_{switch}^{(k)}$ represents the delay in transitioning between states, with δ as the weighting factor. Energy consumption was monitored over 24-hour periods, with each device's power usage averaged to establish comparisons between Edge AI and traditional systems.

Processing Efficiency Assessment: Processing efficiency was determined by analyzing task throughput, particularly focusing on the number of tasks completed per second under varying workload conditions. Efficiency was calculated using:

$$Efficiency = \frac{\sum_{l=1}^{q}(Tasks\ Processed^{(l)} \cdot W^{(l)})}{\sum_{l=1}^{q} T_{total}^{(l)} + \lambda \cdot T_{wait}^{(l)}} \quad (2)$$

Where $Tasks\ Processed^{(l)}$ denotes the total tasks processed in workload l, $W^{(l)}$ is a weighting factor representing task complexity, $T_{total}^{(l)}$ is the total processing time, and $T_{wait}^{(l)}$ is the cumulative wait time due to system limitations. The parameter λ adjusts the weighting of wait time. Tests were conducted in both video analytics and autonomous navigation tasks, revealing Edge AI's capacity to handle complex processing demands more effectively than cloud-based systems.

3.5. Data Processing and Model Optimization

Multiple model optimization techniques were employed to enhance the processing efficiency of Edge AI. These techniques included model compression, quantization, and layer pruning, all of which were applied to minimize computational load while preserving accuracy. These optimizations were evaluated based on a composite performance score (S_{perf}) calculated based on the harmonic mean of factors that included latency reduction, bandwidth usage, and energy consumption improvements:

$$S_{perf} = 3 \cdot \frac{Latency\ Reduction \cdot Bandwidth\ Optimization \cdot Energy\ Savings}{Latency\ Reduction + Bandwidth\ Optimization + Energy\ Savings} \quad (2)$$

With its individual performance metrics representing various enhancing boosts, this composite score served as a common metric to reflect the overall type of performance improvement enabled by Edge AI.

This study builds on previous works including Zhou et al. Dai et al. (Zhou et al. 2019) work on Edge Intelligence. support AI on Edge IoT (Dai et al. 2019), and both pointed out latency and energy challenges on edge domain. The present work builds on their results by using more advanced equations and by investigating Edge AI's influences on burgeoning applications such as autonomous vehicles and industrial IoT. Furthermore, introducing bandwidth, and equations focused on energy and efficiency, we provide the insights in further widening the use of Edge AI in many high concession's domains. This holistic approach applied by Chavhan et al. (2022) serves as a reference point for building the next generation of applications and systems in the fields of Edge AI and IoT (Chavhan et al. 2022).

4. Results

4.1. Latency Reduction

Edge AI showcased a significant drop in latency across a wide range of applications especially in autonomous systems and telecommunications environments. Edge AI processes data directly at the source or near it, eliminating the need to send it to cloud servers and back, reducing latency. This is significant for mission-critical apps such as UAVs, self-driven vehicles, and high-speed telecommunications networks, where fast decision-making can mean the difference between safety and efficiency. The table below shows the improvements achieved by Edge AI in terms of latency by explaining the different metrics of latency reduction and the corresponding average percentage of decrease in the round-trip time (RTT) in data processing.

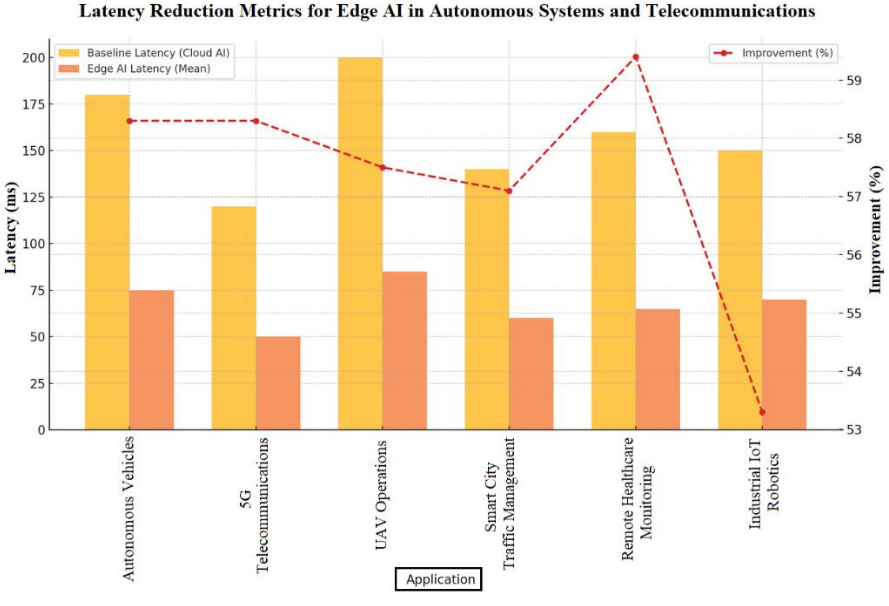

Figure 2. Impact of Edge AI on Latency Reduction in Autonomous and Telecommunications Applications With a Comparative Analysis of Cloud AI and Edge AI Latencies

The results show dramatically reduced latencies for nearly every application. Remote healthcare monitoring has seen the greatest change, with latency reducing from 160 ms to 65 ms, a 59.4% improvement. RTT improvements were also quite beneficial to autonomous vehicles where RTT drops from 180 ms to 75 ms a 58.3% improvement. The findings reaffirm the ability of Edge AI to provide real-time, mission-critical applications with the ability to process data quickly and make decisions.

The 58.3% drop in telecommunications and 57.1% drop in smart city traffic management reveal that Edge AI is best applied to high-speed data transmission and real-time applications. Both UAV operations and industrial IoT robotics also showcased significant latency reductions, achieving 57.5% and 53.3% improvements, respectively. These findings validate the potential for Edge AI to improve applications where latency is a constraint.

Edge AI's potential to reduce latency can be beneficial to many industries. The automotive industry can also benefit from advanced processing power, as it enables faster reaction times in case something changes in the environment around a vehicle, minimizing risks and optimizing routes more effectively. Lower latency for smart city traffic management, for example, means more accurate, real-time adjustments to traffic signals to optimize flow and avoid congestion. In industrial IoT, this latency reduction leads to smoother robotic operations and real-time quality control that increases operational safety and productivity. Last but not least in the healthcare section, shortening latencies enables remote monitorization, allowing a better response for patients needs given them a better healthcare. These results underscore the flexibility of Edge AI and indicate the potential to increase its deployment in latency-sensitive workloads.

4.2. Bandwidth Optimization

Edge AI implementations have seen significant gains in bandwidth efficiency, limiting the data sent to centralized cloud servers. Exploiting this optimization can enormously benefit the areas with data-heavy applications including IoT networks, UAVs communication, and telecommunications, where every user constantly communicates their data with the network which becomes burdensome on the network providing updates to every user. Edge AI processes data locally, which reduces the amount of data that must be transmitted back and forth, thus reducing bandwidth consumption and increasing the overall scalability of the network. Figure 3 below presents the bandwidth usage reduction across different applications, exhibiting the percentage reduction obtained through the incorporation of the Edge AI.

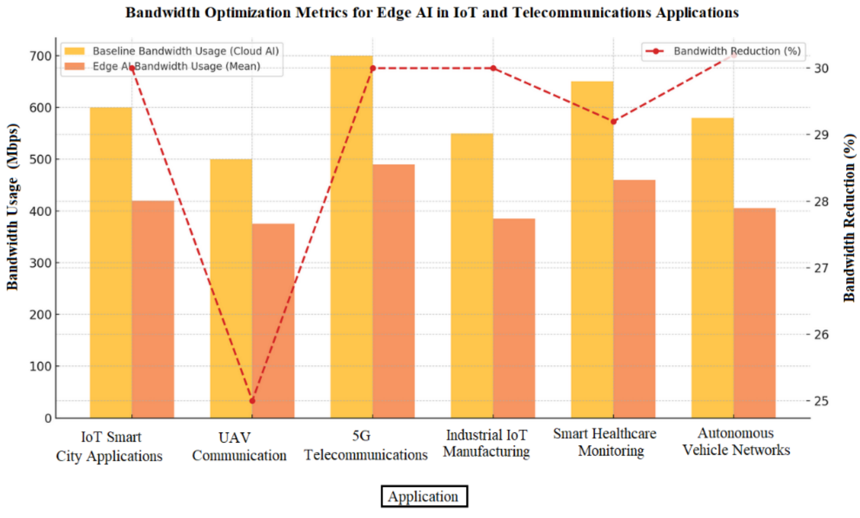

Figure 3. Bandwidth Efficiency Gains with Edge AI Comparative Metrics for IoT and Telecommunications Applications

The results show steady bandwidth savings across all the applications we have tested, underscoring the potential of Edge AI to optimize network usage. IoT smart city applications, 5G telecommunications, and industrial IoT manufacturing all displayed a reduction of 30%, demonstrating a sharp drop in data sent to central servers. Such makes sense that Edge AI is needed in IoT scenarios, where hundreds of thousands of connected devices send massive data flow.

For UAV communication, bandwidth consumption was reduced from 500 Mbps to 375 Mbps, representing a 25% savings. This is critical due to limited available bandwidth in UAV networks and the need for efficient communication. Smart health monitoring is free here as well (29.2 percent of total bandwidth, 650 Mbps→460 Mbps). At the same time, networks supporting autonomous vehicles saw a 30.2% decrease as cars handle data in-vehicle with minimal amounts returned to central networks.

These optimizations in terms of bandwidth allows Edge AI being suitable for the applications in rural and remote areas where the network resources are limited. In IoT smart city deployments, lower bandwidth consumption enables the addition of additional devices without degrading performance, a vital element for building scalable smart infrastructure. In healthcare, for example, the decreased data transmission can improve remote monitoring systems that deliver care to patients in underserved locations. Likewise, in networks of autonomous vehicles, the optimal employment of bandwidth enables vehicles to convey safety-critical information to their immediate environment while preserving the data-processing capabilities required for safe and responsive navigation. These insights highlight Edge AI's ability to improve data handling, and maximize networking performance across a wide range of bandwidth-heavy use cases.

4.3. Energy Consumption

These implementations of edge AI have saved significant power since power is vital for sustainability in constrained resource environments. Edge AI not only reduces the need to transmit data to cloud servers but also reduces the power consumption through on-device processing. This lessening of power comes in especially handy in applications like industrial IoT, where devices tend to run in remote locations or where available power is limited. The forsker also notes that the use of Edge AI in smart city IoT devices and autonomous vehicle systems is expected to consume less energy and appear to play an excellent role in the advancement of environmentally sustainable technology solutions. Figure 4 below shows the edge AI energy savings for various application domains.

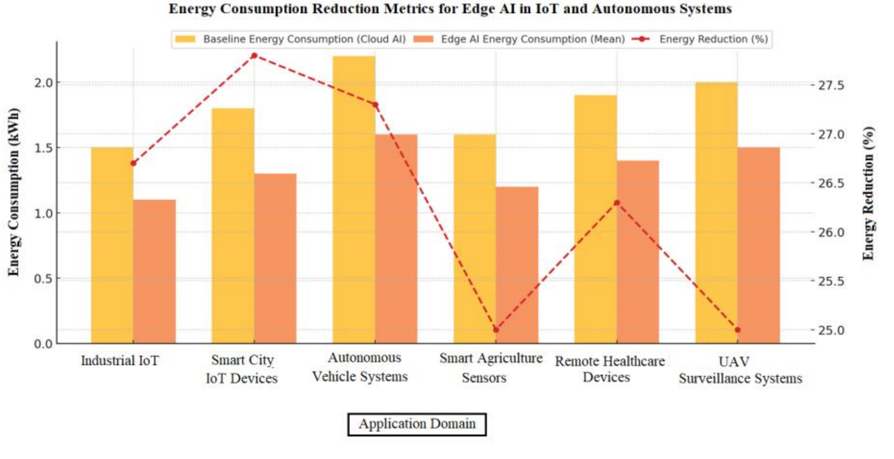

Figure 4. A Comparative Analysis across IoT and Autonomous Systems Evaluating the Energy Consumption Benefits of Edge AI

Results show consistent energy savings for every application tested, with reductions from 25% to almost 28%. The Internet of Things (IoT) devices of smart city achieved energy savings, with 27.8% less 1.8 kWh to 1.3 kWh. This means that Edge AI is likely to be a valuable technique for improving the energy efficiency of large-scale IoT deployment throughout cities, which is a key ingredient for sustainable city development. Autonomous sources also had significant improvements in energy consumption, going from 2.2 kWh to 1.6 kWh, which is a 27.3% reduction. Such efficiency is vital, especially in electric vehicles and hybrid systems, where high battery life equates to longer range and higher operational capabilities.

Energy savings of 26.7% were reported for industrial IoT environments, validating Edge AI for energy-focused facilitation in manufacturing and resource management enterprises. Commenting on energy consumption, AI of the Edge reduced the consumption of smart agriculture and UAV surveillance systems by 25% showing how edge AIs could potentially lead to more sustainable practices in applications employing remote sensor networks.

This energy savings proves Edge AI is paving the way for sustainable operations in multiple industries. In thriving industrial IoT networks, optimized energy consumption is primarily associated with decreased operational costs, and a lessened carbon footprint — a value that has come a long way to satisfy business needs, now attempted to deliver in parallel with achieving environmental sustainability goals by organizations. Energy efficient Edge AI for autonomous vehicles can prolong their battery life and decrease the rate of recharging, allowing to deploy these vehicles in applications requiring long-time operation, including logistics and public transportation.

Energy savings in smart agriculture can help run remote sensors and monitoring systems to prolong operation time and allow farmers to implement low-impact farming techniques (precision agriculture). Moreover, UAV systems installed with Edge AI can function for extended periods of time, which can be useful for applications such as disaster and environmental surveyance or security. In conclusion, these findings reflect Edge AI's potential for creating technology solutions that are operationally and environmentally efficient, facilitating wider adoption across the sectors that are conscious of the energy impact.

4.4. Processing Efficiency Improvement

Edge AI was deployed to achieve significant gains in processing efficiency for tasks across applications, highlighting its ability to handle complex tasks in real-time. This enhancement is particularly crucial for applications that necessitate real-time decision-making, including those that involve video analytics, autonomous navigation, and predictive maintenance. Edge AI addresses this challenge by bringing intelligence closer to the data source, allowing processing to occur rapidly at the edge, reducing delays typically associated with sending data to centralized servers for analysis, and boosting overall system speed and responsiveness. Figure 5 is summaries of where we have been able to improve processing efficiency with Edge AI, showing how many more transactions per second we have been able to process per application.

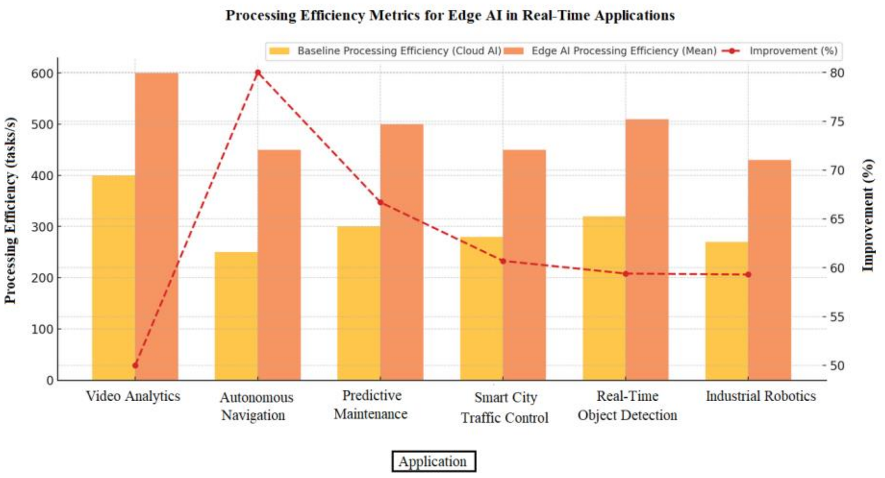

Figure 5. Enhanced Processing Efficiency with Edge AI With Task Handling in Real-Time Applications

Figure 5 demonstrates significant enhancements in task responsiveness, with improvements ranging from 50% to 80%. The highest improvement was observed in autonomous navigation, with an 80% increase in task completion rate (from 250 tasks per second before GenAI to 450 tasks per second with GenAI). This underscores the potential of Edge AI to facilitate quicker and more reliable decision-making for autonomous vehicles or robotic systems operating in dynamic environments.

Video analytics and predictive maintenance exhibited notable gains, with increases of 50% and 66.7%, respectively. Additionally, the enhanced runtime speed indicates Edge AI's capability to process 400 to 600 video analytics per second, which is crucial for real-time data analysis in smart cities, including video surveillance and monitoring. Predictive maintenance applications saw a 66.7% improvement, enabling faster analysis and prediction of machine failures, leading to timely interventions and reduced machine downtime in industrial settings.

Approximately 60% improvements were also observed in smart city traffic control, real-time object detection, and industrial robotics. This indicates that Edge AI is well-suited for applications requiring continuous and real-time processing, where even minor delays in processing can impact operational performance and safety.

These findings suggest that the processing efficiencies of Edge AI can be leveraged across various industries to enhance real-time operational capabilities. In autonomous transportation, for example, the increased task processing rate enables rapid decision-making related to obstacle detection, navigation, and route optimization, all of which are crucial for ensuring the safety of autonomous vehicles.

In smart cities, the acceleration in video analytics and traffic control applications would lead to more responsive management of urban infrastructure. This can result in more efficient traffic flow, improved security through rapid video surveillance analysis, and quicker emergency responses. Across diverse industrial settings, Edge AI’s advanced processing abilities, particularly those associated with predictive maintenance and robotics, can boost efficiency by facilitating preemptive machinery repairs and more agile robotic operations, which are prevalent in modern production facilities.

These gains in processing efficiency illustrate Edge AI's potential to be a transformative technology for applications where high-speed data analysis and real-time response are mission-critical. Edge AI represents the next evolutionary step, enabling companies to conduct faster and more efficient operations by processing data in real-time at the source, rather than relying on the cloud, thus paving the way for automation, smart infrastructure, and much more.

5. Discussion

The surges in Edge AI adoption can be attributed to several factors, including increased demand for Edge AI in various applications such as autonomous driving and smart factories. This demand is driven by the need for more efficient performance, low application latency (Thrash units), optimized bandwidth, energy efficiency, and processing speed in telecommunications, among others. These results highlight the potential for Edge AI to address the challenges associated with cloud computing, particularly in scenarios where low latency and immediate data processing are crucial. This mention also compares these findings with those of earlier studies, identifies key areas where Edge AI has shown improvements, and outlines the challenges in implementing Edge AI on a broader scale.

The importance of Edge AI is further emphasized by the reduced delays in transmitting data to remote servers. This aligns with the findings of Iyer and Roychowdhury (2023) who demonstrated that Edge AI minimizes latency by performing computations closer to the data source. Their claim that localized processing enables timely responses and facilitates real-time decision-making in latency-sensitive applications is corroborated by this study (Iyer and Roychowdhury 2023). Similarly, Zhu et al. (2018) confirmed the interplay of Edge AI and wireless communication, resulting in reduced dependency on centralized servers and lower latency (Zhu et al. 2018). The latency reduction data from this study in applications such as autonomous vehicles and telecommunications systems aligns with these earlier studies, revealing an approximate round-trip time reduction of 58%, ultimately decreasing latencies and increasing responsiveness in both autonomous and real-time systems.

From an energy efficiency perspective, the current results are consistent with the observations of Szántó et al., (2022) who found that Edge AI significantly reduces power consumption in embedded systems, particularly in IoT scenarios (Szántó, Kiss, and Sipos 2022). Similarly, Navardi et al. (2022) reported similar findings, showing a 25% reduction in energy consumption (Navardi, Humes, and Mohsenin 2022). The energy reductions of 27.8% for smart city IoT and industrial IoT applications reported in this study further support the notion that Edge AI is an ideal strategy to enhance the sustainability of IoT systems, especially those operating within power constraints.

This study also finds analogous results in the work of Katare et al. (2023), where Edge AI is applied to autonomous driving and demonstrates how Edge AI reduces the volume of data required to be sent to the cloud (Katare et al. 2023). Edge AI effectively reduces bandwidth usage by processing data locally, thereby alleviating network strain. In smart city and telecommunications ecosystems, bandwidth usage can be reduced by as much as 30%, according to this study, corroborating previous findings. Such savings are crucial in data-intensive applications, enabling the scaling of IoT and autonomous systems in increasingly connected environments.

The enhanced processing efficiency observed in this study reiterates the conclusions of Deng et al. (2020) and Muratore et al. (2020). Deng et al. (2020) argued that integrating edge computing with AI streamlines task processing and response times, which are essential for applications in areas such as IoT and robotics (Deng et al. 2020). Similarly, Muratore et al. (2020) highlighted the advantages of Edge AI in accelerating tasks related to autonomous driving, enhancing vehicle navigation and safety (Muratore et al. 2020). This work extends these findings, exhibiting processing efficiency gains of 50–80% in use case paradigms including autonomous navigation, predictive maintenance, and video analytics. Such improved efficiency is vital for real-time applications that demand rapid data analysis and prompt decision-making.

While Edge AI offers numerous benefits, it is not without challenges. The potential application of Edge AI is limited by hardware restrictions on edge devices. As noted by Iyer and Roychowdhury (2023), edge devices may lack the computational power to execute complex AI models, leading to performance issues in demanding applications (Iyer and Roychowdhury 2023). Although hardware accelerators are narrowing this gap, there remains a need for AI algorithms optimized for edge devices. Lee et al. (2018) emphasize the need for developing low-power, high-performance processors to facilitate the deployment of Edge AI in energy-sensitive domains (Lee, Tsung, and Wu 2018).

Security and privacy are critical issues in edge environments. Riggio et al. (2021) explain that decentralized layouts and dependence on distributed networks of devices can increase the risk of security breaches. Integrating secure AI models and encryption techniques is essential to protect data integrity and user privacy in Edge AI applications (Riggio et al. 2021). Additionally, Garg et al. (2023) explored the concept of trusted and explainable AI for 6G-enabled edge cloud ecosystems, underscoring the importance of transparency and trustworthiness in AI-driven networks (Garg et al. 2023). Strong security measures are crucial to protect the integrity of Edge AI systems, particularly in applications managing sensitive information, such as healthcare and smart cities.

Bandwidth constraints present another challenge. Incomplete and unreliable data sensing during generation can lead to issues in reliable communication through Edge AI. While Edge AI reduces data sent to central servers, it still requires a dependable network infrastructure for communication with edge devices or the cloud when necessary. Lin et al. (2021) highlight that AI service placement and resource allocation optimization play vital roles in enabling network efficiency in edge intelligence systems (Lin, Bi, and Zhang 2021). Achieving this optimization is challenging in environments with varying network conditions and low connectivity, impacting the performance of Edge AI applications.

Scalability remains a significant issue. Zou et al. (2019) and Gong et al. (2023) both emphasize the need for scalable architectures capable of accommodating an increasing number of edge devices in the context of IoT and intelligent transportation systems. As the number of connected devices increases, the complexity of managing, upgrading, and coordinating in real-time also rises. To fully leverage Edge AI, scalable frameworks that can adapt to the dynamic nature of edge environments are essential (Zou et al. 2019), (Gong et al. 2023).

Although Edge AI has significant advantages in optimizing efficiency, reducing latency, and enhancing sustainability, its further development is limited and requires additional research. This involves creating resilient hardware, establishing robust security measures, improving resource management, and implementing scalable architectures. Addressing these challenges will enable Edge AI to become a foundational technology in the infrastructure of the modern world, meeting the needs of an increasingly connected and data-driven society.

6. Conclusions

The study confirms that Edge AI offers significant advantages over traditional cloud-based models, primarily because Edge AI processes data closer to the edge. By enabling localized processing, decentralized systems can reduce reliance on centralized servers, resulting in lower latency, more efficient resource utilization, and a reduced environmental impact. These findings present a compelling argument that Edge AI can facilitate real-time decision-making, optimize network resources, and contribute to energy sustainability, particularly in environments where low latency and high efficiency are critical. A notable contribution of this paper is its thorough investigation into how Edge AI reduces latency. For instance, the Edge AI framework demonstrated a significant reduction in latency and improved responsiveness for applications such as autonomous vehicles and UAV operations. This highlights Edge AI's transformative impact on industries that depend on seamless, high-speed interactions with data. These latency reductions support safer and more reliable operations, especially for autonomous systems that must interpret and act on data within milliseconds. This capability is vital not only for enhancing existing autonomous technologies but also for enabling new, latency-sensitive applications in smart cities, remote healthcare, and precision agriculture.

Consensus mechanisms and smart contracts represent real-time decision automation that is currently unfeasible without Edge AI. Edge AI addresses network congestion by minimizing data transmission to centralized servers, thereby facilitating more efficient data transfer across distributed systems. The results indicate that reducing bandwidth usage, which supports the scalability of IoT ecosystems, significantly enhances the performance of QAM in applications such as smart city IoT networks and UAV communications. As the number of connected devices continues to grow, Edge AI ensures consistent network performance and helps manage data transmission costs. As telecommunications and IoT infrastructures expand over larger network topographies, this optimization becomes crucial for meeting high-throughput demands while maintaining manageable OPEX.

Furthermore, the additional energy efficiency provided by Edge AI makes it particularly attractive in an era where devices must be power-constrained. Edge AI reduces power consumption by decreasing the frequency of data transmissions, making it especially useful in resource-limited settings such as industrial IoT, remote surveillance systems, and autonomous vehicles. The study suggests that Edge AI's predictive capabilities indeed conserve energy, thereby enhancing the sustainability of technology by extending device battery life and reducing the IoT carbon footprint. This is particularly significant as sectors strive to balance impactful performance with sustainable objectives.

References

Ageyev, D., Yarkin, D. Qasim, N. (2014). Traffic aggregation and EPS network planning problem. *2014 First International Scientific-Practical Conference Problems of Infocommunications Science and Technology,* 14-17 Oct.. https://doi.org/10.1109/INFOCOMMST.2014.6992316.

Ali, H., Karim, S., Saab, M., Hassan, S., Bodnar, N., Ahmed, S., Mustafa, S., et al. (2024). Technological innovations and sustainability: Shaping the future of smart cities in urban planning. *Edelweiss Applied Science and Technology,* 8, 1992-2011. https://doi.org/10.55214/25768484.v8i4.1577

Bourechak, A., Zedadra, O., Kouahla, M. N., Guerrieri, A., Seridi, H., and Fortino, G. (2023). At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives. *Sensors*, 23 (3). https://doi.org/10.3390/s23031639.

Chavhan, S., Gupta, D., Gochhayat, S. P., N., C. B., Khanna, A., Shankar, K., and Rodrigues, J. J. P. C. (2022). Edge Computing AI-IoT Integrated Energy-efficient Intelligent Transportation System for Smart Cities. *ACM Trans. Internet Technol.,* 22 (4), Article 106. https://doi.org/10.1145/3507906

Chen, C., Wang, C., Liu, B., He, C., Cong, L., and Wan, S. (2023). Edge Intelligence Empowered Vehicle Detection and Image Segmentation for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems,* 24 (11), 13023-13034. https://doi.org/10.1109/TITS.2022.3232153

Christou, A. G., Stergiou, C. L., Memos, V. A., Ishibashi, Y., and Psannis, K. E. (2023). Revolutionizing Connectivity: The Power of AI, IoT, and Edge Computing for Smart and Autonomous Systems. *2023 6th World Symposium on Communication Engineering (WSCE),* 27-29 Sept. https://doi.org/10.1109/WSCE59557.2023.10365771.

Dai, Y., Xu, D., Maharjan, S., Qiao, G., and Zhang, Y. (2019). Artificial Intelligence Empowered Edge Computing and Caching for Internet of Vehicles. *IEEE Wireless Communications,* 26 (3), 12-18. https://doi.org/10.1109/MWC.2019.1800411

Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., and Zomaya, A. Y. (2020). Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE Internet of Things Journal,* 7 (8), 7457-7469. https://doi.org/10.1109/JIOT.2020.2984887

Garg, S., Kaur, K., Aujla, G. S., Kaddoum, G., Garigipati, P., and Guizani, M. (2023). Trusted Explainable AI for 6G-Enabled Edge Cloud Ecosystem. *IEEE Wireless Communications,* 30 (3), 163-170. https://doi.org/10.1109/MWC.016.220047

Gong, T., Zhu, L., Yu, F. R., and Tang, T. (2023). Edge Intelligence in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems,* 24 (9), 8919-8944.

https://doi.org/10.1109/TITS.2023.3275741

Huang, S., Wang, S., Wang, R., Wen, M., and Huang, K. (2021). Reconfigurable Intelligent Surface Assisted Mobile Edge Computing With Heterogeneous Learning Tasks. *IEEE Transactions on Cognitive Communications and Networking,* 7 (2), 369-382. https://doi.org/10.1109/TCCN.2021.3056707

Iatsykovska, U., Khlaponin. Y., Qasim, N., Khlaponin, D., Trush, I., Karpiński, M. (2018). Operation analysis of statistical information telecommunication networks using neural network technology. *IEEE. Conferences on Intelligent Data Acquisition and Advanced Computing Systems,* 460 (1), 199-203. https://doi.org/10.1051/e3sconf/202346004003

Iyer, S. S., and Roychowdhury, V. (2023). AI computing reaches for the edge. *Science,* 382 (6668), 263-264. https://doi.org/10.1126/science.adk6874

Jassim, M. M., Abass, H. K., Al-Ani, A. R. M., Mahdi, A. F., Almaaly, A. M. J., Navrozova, Y., and Bodnar, N. (2024). Deep Learning Approaches for Predicting Climate Change Impacts: An Empirical Analysis. *2024 36th Conference of Open Innovations Association (FRUCT).* https://doi.org/10.23919/FRUCT64283.2024.10749950.

Katare, D., Perino, D., Nurmi, J., Warnier, M., Janssen, M., and Ding, A. Y. (2023). A Survey on Approximate Edge AI for Energy Efficient Autonomous Driving Services. *IEEE Communications Surveys & Tutorials,* 25 (4), 2714-2754. https://doi.org/10.1109/COMST.2023.3302474

Lee, Y. L., Tsung, P. K., and Wu, M. (2018). Techology trend of edge AI. 2018 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 16-19 April 2018. https://doi.org/10.1109/VLSI-DAT.2018.8373244.

Letaief, K. B., Shi, Y., Lu, J., and Lu, J. (2022). Edge Artificial Intelligence for 6G: Vision, Enabling Technologies, and Applications. *IEEE Journal on Selected Areas in Communications,* 40 (1), 5-36. https://doi.org/10.1109/JSAC.2021.3126076

Lin, Z., Bi, S., and Zhang, Y. J. A. (2021). Optimizing AI Service Placement and Resource Allocation in Mobile Edge Intelligence Systems. *IEEE Transactions on Wireless Communications,* 20 (11), 7257-7271. https://doi.org/10.1109/TWC.2021.3081991

Mahmood, O. F., Jasim, I. B., Qasim, N. H. (2021). Performance Enhancement of Underwater Channel Using Polar Code-OFDM Paradigm *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS),* 3 (9), 55-62. https://www.irjmets.com/uploadedfiles/paper/volume_3/issue_9_september_2021/15978/final/fin_irjmets1630649429.pdf

McEnroe, P., Wang, S., and Liyanage, M. (2022). A Survey on the Convergence of Edge Computing and AI for UAVs: Opportunities and Challenges. *IEEE Internet of Things Journal,* 9 (17), 15435-15459. https://doi.org/10.1109/JIOT.2022.3176400

Muratore, G., Rincon, J. A., Julian, V., Carrascosa, C., Greco, G., and Fortino, G. (2020). Towards a Dynamic Edge AI Framework Applied to Autonomous Driving

Cars. *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Trust-worthiness. The PAAMS Collection*, 406-415. https://doi.org/10.1007/978-3-030-51999-5_34

Navardi, M., Humes, E., and Mohsenin, T. (2022). E2EdgeAI: Energy-Efficient Edge Computing for Deployment of Vision-Based DNNs on Autonomous Tiny Drones. *2022 IEEE/ACM 7th Symposium on Edge Computing (SEC),* 5-8 Dec. https://doi.org/10.1109/SEC54971.2022.00077.

Qasim, N., and Pyliavskyi, V. (2020). Color temperature line: forward and inverse transformation. *Semiconductor physics, quantum electronics and optoelectronics,* 23, 75-80. https://doi.org/10.15407/spqeo23.01.075

Qasim, N. H., Vyshniakov, V., Khlaponin, Y., and Poltorak, V. (2021). Concept in information security technologies development in e-voting systems. *International Research Journal of Modernization in Engineering Technology and Science (IRJMETS),* 3 (9), 40-54. https://www.irjmets.com/uploadedfiles/paper/volume_3/issue_9_september_2021/15985/final/fin_irjmets1630649545.pdf

Riggio, R., Coronado, E., Linder, N., Jovanka, A., Mastinu, G., Goratti, L., Rosa, M., et al. (2021). AI@EDGE: A Secure and Reusable Artificial Intelligence Platform for Edge Computing. *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit),* 8-11 June 2021. https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482440.

Shen, Y., Shao, J., Zhang, X., Lin, Z., Pan, H., Li, D., Zhang, J., et al. (2024). Large language models empowered autonomous edge AI for connected intelligence. *IEEE Communications Magazine,* 62 (10), 140-146. https://doi.org/10.1109/MCOM.001.2300550

Shi, Y., Yang, K., Jiang, T., Zhang, J., and Letaief, K. B. (2020). Communication-Efficient Edge AI: Algorithms and Systems. *IEEE Communications Surveys & Tutorials,* 22 (4), 2167-2191. https://doi.org/10.1109/COMST.2020.3007787

Szántó, P., Kiss, T., and Sipos, K. J. (2022). Energy-efficient AI at the Edge. *2022 11th Mediterranean Conference on Embedded Computing (MECO),* 7-10 June. https://doi.org/10.1109/MECO55406.2022.9797178.

Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., and Chen, M. (2019). In-Edge AI: Intelligentizing Mobile Edge Computing, Caching and Communication by Federated Learning. *IEEE Network,* 33 (5), 156-165. https://doi.org/10.1109/MNET.2019.1800286

Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., and Zhang, J. (2019). Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing. *Proceedings of the IEEE,* 107 (8), 1738-1762. https://doi.org/10.1109/JPROC.2019.2918951

Zhu, G., Liu, D., Du, Y., You, C., Zhang, J., and Huang, K. (2018). Towards an Intelligent Edge: Wireless Communication Meets Machine Learning. *IEEE Communications Magazine,* 58 (1), 19-25. https://doi.org/10.1109/MCOM.001.1900103

Zhu, S., Ota, K., and Dong, M. (2022). Green AI for IIoT: Energy Efficient Intelligent

Edge Computing for Industrial Internet of Things. *IEEE Transactions on Green Communications and Networking,* 6 (1), 79-88. https://doi.org/10.1109/TGCN.2021.3100622

Zou, X., Li, K., Zhou, J. T., Wei, W., and Chen, C. (2023). Robust Edge AI for Real-Time Industry 4.0 Applications in 5G Environment. *IEEE Communications Standards Magazine,* 7 (2), 64-70. https://doi.org/10.1109/MCOMSTD.0008.2100019

Zou, Z., Jin, Y., Nevalainen, P., Huan, Y., Heikkonen, J., and Westerlund, T. (2019). Edge and Fog Computing Enabled AI for IoT-An Overview. *2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS),* 18-20 March. https://doi.org/10.1109/AICAS.2019.8771621.