



Paper Type: Original Article

AI-Enhanced Data Processing in Cloud-Driven IoT Systems

Ridwan Boya Marqas^{1,*} , Shaikh Saquib Jamal², Ankur Singh²

¹ Department of Computer Science, College of Science, Knowledge University, Erbil, Iraq; ridwan.marqas@dpu.edu.krd.

² School of Computer Science Engineering, KIIT University, Bhubaneswar, India; 23057046@kiit.ac.in; 2205970@kiit.ac.in.

Citation:

Received: 28 August 2025

Revised: 30 October 2025

Accepted: 13 January 2026

Marqas, R. B., Jamal, S. S., & Singh, A. (2026). AI-enhanced data processing in cloud-driven IoT systems. *Smart Internet of Things*, 3(3), 83-91.

Abstract

The convergence of Artificial Intelligence (AI), cloud computing, and Internet of Things (IoT) systems has led to significant advancements in how large-scale data is processed, managed, and utilized. AI provides advanced analytic capabilities, including Machine Learning (ML), predictive analytics, and real-time decision-making, which are critical for handling the massive datasets generated by IoT devices. Cloud computing platforms offer scalable infrastructure that integrates seamlessly with AI technologies, enabling enhanced IoT data processing. This paper explores how AI improves cloud-driven IoT systems, addressing key challenges such as scalability, data privacy, and energy efficiency. The study delves into the architecture of AI-enhanced cloud platforms, examines AI algorithms used in IoT data processing, and highlights the security challenges associated with integrating AI in cloud-driven IoT systems. Emerging trends such as federated learning, AIoT, and the impact of 5G networks on AI-enhanced IoT systems are also discussed. The paper concludes by identifying future research directions and practical implications for industries seeking to adopt AI-driven IoT solutions.

Keywords: Artificial intelligence, Internet of things, Cloud computing, Data processing, Machine learning, AIoT, Federated learning, 5G.

1 | Introduction

The rapid proliferation of Internet of Things (IoT) devices has dramatically altered the landscape of data generation and consumption. IoT systems encompass interconnected devices such as sensors, cameras, wearables, and industrial machines that collect and transmit data in real time. These devices are used across industries including healthcare, manufacturing, agriculture, transportation, and smart cities, contributing to what is estimated to be more than 50 billion connected devices by 2030.

However, managing and processing the massive amounts of data generated by these IoT devices presents significant challenges. Traditional data processing architectures are often insufficient to handle the volume,

 Corresponding Author: ridwan.marqas@dpu.edu.krd

 <https://doi.org/10.48313/siot.v3i2.89>



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

velocity, and variety of IoT data. The increasing complexity of data demands more than just storage capabilities; it requires robust analytics to extract actionable insights in real time.

Cloud computing has become a key enabler of IoT by providing scalable, on-demand resources for data storage and processing. By offloading the heavy computational tasks to cloud platforms, IoT systems benefit from virtually unlimited storage and processing power. However, while cloud computing addresses the challenges of scalability and data storage, the integration of Artificial Intelligence (AI) is revolutionizing how cloud-driven IoT systems handle data. AI algorithms, including Machine Learning (ML) and Deep Learning (DL) models, offer the capability to process IoT data intelligently, enabling predictions, automated decisions, and real-time optimizations.

At present, there are four major challenges: First, limited computing resources, and the computing power of Edge computing devices is generally lower than that of cloud servers. Therefore, in combination with the complexity and scale of AI models, Edge computing has a certain phenomenon of hardware accelerator weakening. Therefore, in the process of improvement, AI algorithms can be implemented on Edge devices. Larger, more efficient models must be developed to enable more efficient data processing. The second is the energy efficiency of Edge computing. It can be seen that ensuring the security of Edge artificial systems and preventing data leakage is one of the biggest research issues of Edge computing. The last challenge is model update and maintenance. Due to the relatively wide distribution of Edge devices, effective device maintenance is a relatively big challenge. Therefore, distributed mechanisms, federated learning, and other technologies need to be combined in daily maintenance. Fig. 1 shows an architecture diagram of the current challenges of Edge computing devices [1].

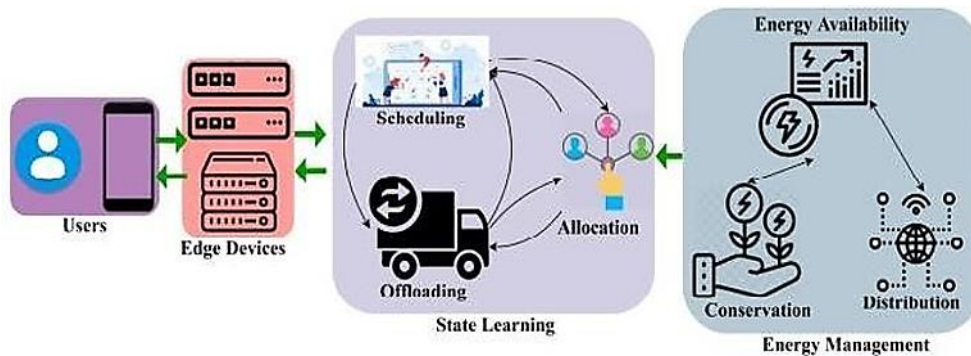


Fig. 1. Architecture diagram of the challenges of edge computing devices.

2| Internet of Things Data Challenges

2.1| The 3V's of Internet of Things Data: Volume, Velocity, and Variety

The most significant challenge of IoT systems is managing the massive amount of data generated by billions of connected devices. These data are typically characterized by the 3Vs [2]:

- I. **Volume:** IoT devices produce enormous quantities of data. For instance, a single autonomous vehicle can generate up to 5 terabytes of data per day from sensors monitoring speed, acceleration, Global Positioning System (GPS), camera footage, radar, and Light Detection and Ranging (LiDAR). In smart cities, sensors monitoring traffic, pollution, weather, and public utilities generate continuous data streams that need to be processed in real time.
- II. **Velocity:** The speed at which IoT devices generate and transmit data is a critical factor. Many applications, such as autonomous vehicles, healthcare monitoring systems, and industrial automation, require real-time processing with near-instantaneous decision-making. Delays in processing can result in catastrophic consequences, such as accidents or system failures.

- I. Variety: IoT data comes in various formats, including structured data (sensor readings), unstructured data (video, audio), and semi-structured data (logs, XML). This diversity in data types complicates the processing, storage, and analysis, requiring versatile and adaptive data processing solutions.

2.2 | Real-time Processing Requirements

Many IoT applications require real-time processing of data to ensure the system's smooth operation. In smart transportation systems, for instance, sensors on roads and traffic lights generate data continuously, and any delay in processing can lead to congestion, accidents, or inefficient energy consumption. Similarly, in healthcare, wearable devices monitor patients' vital signs in real time, alerting doctors to any abnormalities that could indicate a medical emergency. The critical nature of these applications demands fast, efficient processing of IoT data streams.

Traditional data processing methods struggle to meet these requirements, particularly in environments where processing needs fluctuate rapidly. AI-enhanced cloud computing platforms, however, are capable of handling the dynamic nature of IoT workloads by employing ML models to predict and optimize processing needs in real time.

2.3 | Resource Constraints

Another major challenge in IoT systems is the limited computational resources of IoT devices themselves. IoT devices, especially those deployed at the edge (e.g., sensors, microcontrollers), typically have limited memory, processing power, and energy resources. These devices rely on centralized cloud computing platforms to handle computationally expensive tasks like data aggregation, pattern recognition, and ML - based prediction.

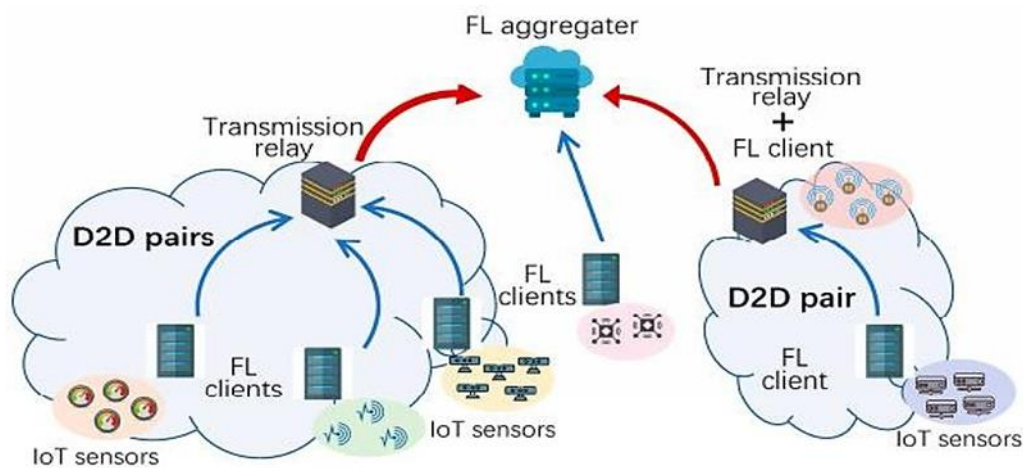


Fig. 2.

However, transmitting large amounts of raw data from IoT devices to cloud platforms can strain network bandwidth and increase latency, particularly in environments where low latency is critical. Edge computing, which involves processing data closer to the source before sending it to the cloud, helps alleviate these issues by reducing the amount of data that needs to be transmitted. However, even with edge computing, the integration of AI into cloud platforms is essential for more advanced analytics and decision-making capabilities [3].

3 | Cloud-Driven Internet of Things Architecture

To address the complexities of IoT data processing, cloud computing architectures are typically employed to handle the vast computational and storage demands. The architecture of cloud-driven IoT systems can be broken down into three layers: Edge Layer, Edge Computing, and Cloud Layer [1], [4].

3.1| Edge Layer

The edge layer comprises the IoT devices themselves, including sensors, cameras, wearables, and industrial machines. These devices collect data from their environments, such as temperature, pressure, motion, video, and sound. Given the limitations of these devices in terms of computational power and storage, they often perform only minimal processing, such as filtering or compressing data, before transmitting it to the cloud.

For example, in a smart agriculture setup, sensors in the field monitor soil moisture levels, temperature, and humidity. These sensors transmit the data to the cloud, where it is analyzed and used to optimize irrigation schedules, reducing water consumption while improving crop yield.

3.2| Edge Computing

To mitigate the latency and bandwidth challenges associated with transmitting all IoT data to centralized cloud servers, edge computing has emerged as a complementary technology. Edge computing refers to the practice of processing data closer to where it is generated, reducing the need to send all data to the cloud. By processing data at the edge, organizations can minimize latency and improve response times for time-sensitive applications.

For instance, in autonomous driving, real-time decisions about braking, steering, and obstacle avoidance cannot wait for data to be transmitted to a cloud server and back. Edge computing nodes, often embedded in the vehicle or roadside units, process the data locally to ensure near-instantaneous decision-making.

Edge computing is particularly beneficial in applications where bandwidth is limited or where real-time processing is critical. However, complex tasks such as DL model training, large-scale data analysis, and long-term storage are still performed in the cloud.

3.3| Cloud Layer

The cloud layer is the backbone of IoT data processing, providing scalable resources for storing, processing, and analyzing the massive amounts of data generated by IoT devices. Cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud offer various services tailored to IoT applications, including:

- I. AI services: These platforms provide AI tools for analyzing IoT data, including pre-trained ML models, neural networks, and real-time analytics services. AI services enable predictive maintenance, anomaly detection, and real-time decision-making for IoT systems.
- II. Data storage: Cloud platforms provide scalable storage solutions, including object storage (for unstructured data such as images and video) and relational databases (for structured data). These systems support the storage of vast amounts of IoT data while ensuring data availability and security.

Cloud platforms also offer services such as serverless computing (e.g., AWS Lambda), which allows IoT applications to scale automatically in response to changes in workload, reducing operational costs and improving efficiency.

4| Artificial Intelligence in Internet of Things Data Processing

AI is a critical enabler of intelligent data processing in cloud-driven IoT systems. By leveraging advanced AI technologies such as ML, DL, and Reinforcement Learning (RL), cloud platforms can analyze IoT data in real time, enabling predictive maintenance, anomaly detection, and automated decision-making. Before we start with IoT, we can explain Data Analysis Models [5].

4.1| Data Analysis Models

We can divide data analytics research into six key technical areas: structured data analysis, text data

analysis, web data analysis, multimedia data analysis, network data analysis, and mobile data analysis. Such a classification aims to emphasize the characteristics of the data, but some areas may use similar basic technologies. Since data analysis has a wide scope and comprehensive coverage is not easy, we will focus on key problems and technologies in data analysis in the following discussions [6].

4.2 | Machine Learning in Internet of Things

ML algorithms are widely used in IoT applications for predictive analytics and decision-making. ML algorithms can learn from historical data generated by IoT devices and make predictions about future events. These predictions can then be used to automate decision-making processes, optimize operations, and reduce costs.

For example, in smart manufacturing, IoT sensors monitor machines in real time, collecting data on temperature, vibration, and operating conditions. ML models analyze this data to predict when a machine is likely to fail, allowing maintenance teams to schedule repairs before a breakdown occurs. This approach, known as predictive maintenance, reduces downtime and minimizes repair costs by preventing equipment failures.

4.3 | Deep Learning for Internet of Things Data

DL, a subset of ML, is particularly useful for processing large, complex datasets such as images, audio, and video. DL models use neural networks with multiple layers to detect patterns in data and make predictions [7]. These models are especially powerful in applications that require advanced pattern recognition, such as image classification, speech recognition, and natural language processing.

In smart cities, for example, DL models are used to analyze video feeds from traffic cameras. The models detect patterns in traffic flow, predict congestion, and suggest optimal routes for vehicles.

Additionally, DL is used in facial recognition systems, enabling cities to enhance public safety by identifying individuals in real-time based on video surveillance footage.

4.4 | Reinforcement Learning for Dynamic Internet of Things Environments

RL is a type of ML where agents learn to make decisions by interacting with their environment. RL is particularly useful in dynamic environments where conditions change over time, such as autonomous systems and industrial automation.

For instance, in autonomous drones, RL models allow the drone to learn from its environment and optimize its flight path in real time. The drone receives feedback from its sensors (e.g., obstacle detection, wind speed) and adjusts its actions to avoid obstacles, minimize energy consumption, and complete its mission efficiently.

RL is also used in smart grid systems, where AI agents learn to balance energy supply and demand by optimizing the distribution of electricity. These systems can automatically adjust energy production based on real-time data from smart meters, improving energy efficiency and reducing costs.

5 | Artificial Intelligence -Enhanced Cloud Platforms

Cloud platforms are increasingly integrating AI services to handle the complex data processing needs of IoT systems. These AI-enhanced platforms provide tools and services that enable developers to build intelligent IoT applications without requiring deep expertise in AI.

5.1 | Artificial Intelligence-as-a-Service

Many cloud providers now offer Artificial Intelligence-as-a-Service (AIaaS) platforms, which allow businesses to leverage AI capabilities without needing to build their own AI infrastructure. AIaaS platforms provide pre-trained ML models, APIs for real-time analytics, and tools for training custom models on IoT data.

For example, AWS offers SageMaker, a fully managed service that allows developers to build, train, and deploy ML models at scale. Google Cloud AI provides AutoML, a suite of ML tools that allow businesses to train high-quality models on their data with minimal expertise. Microsoft Azure AI offers similar services, including Azure ML, which enables businesses to build, train, and deploy AI models on cloud-based IoT data.

These platforms enable businesses to build intelligent IoT applications quickly and efficiently, reducing the complexity of integrating AI into IoT systems.

5.2 | Distributed Artificial Intelligence Models and Federated Learning

A major trend in AI-enhanced cloud platforms is the development of distributed AI models and federated learning. Federated learning is a decentralized approach to AI model training, where models are trained locally on IoT devices, and only the model updates are sent to the cloud. This decentralized training approach reduces the amount of data that needs to be transmitted, improving privacy and reducing bandwidth usage.

Federated learning is particularly useful in applications where data privacy is critical, such as healthcare. For example, wearable devices can collect health data from patients and train AI models locally without transmitting sensitive data to the cloud. The model updates are then aggregated in the cloud, allowing healthcare providers to make predictions without compromising patient privacy [8].

Federated learning is also being used in autonomous vehicles, where AI models are trained locally on data collected by the vehicle's sensors. These models are updated in real-time based on the vehicle's interactions with its environment, improving the vehicle's decision-making capabilities without relying on constant communication with the cloud.

6 | Data Security and Privacy Concerns

The integration of AI in cloud-driven IoT systems presents significant security and privacy challenges. IoT devices collect vast amounts of sensitive data, including personal health information, location data, and financial transactions. Ensuring the security and privacy of this data is critical, particularly when AI is involved in processing and decision-making.

6.1 | Data Integrity and Trust

One of the key challenges in AI-enhanced IoT systems is ensuring the integrity of the data being processed. If the data is corrupted or tampered with, AI models may produce incorrect predictions, leading to potentially harmful outcomes. For example, in healthcare, corrupted data from wearable devices could lead to incorrect diagnoses or treatment recommendations.

To address this challenge, cloud platforms must implement strong data validation and integrity-checking mechanisms to ensure that the data being used by AI models is accurate and trustworthy. Blockchain technology is also being explored as a potential solution for ensuring data integrity in IoT systems, as it provides a tamper-proof record of all data transactions.

6.2 | Privacy Concerns in Artificial Intelligence-Driven Internet of Things Systems

Privacy is a major concern in IoT systems, particularly in sectors like healthcare and finance, where sensitive personal data is collected and processed. AI models rely on large amounts of data to make accurate predictions, but transmitting sensitive data to the cloud for analysis raises concerns about privacy and data protection.

Data anonymization techniques are commonly used to protect sensitive data in AI-driven IoT systems. These techniques remove Personally Identifiable Information (PII) from the data, allowing AI models to analyze the data without compromising user privacy. However, anonymization is not foolproof, and re-identification attacks (where anonymized data is combined with other datasets to reveal PII) remain a concern [9], [10].

Federated learning, as discussed earlier, is another approach to improving privacy in AI-enhanced IoT systems. By keeping data local and only sharing model updates, federated learning reduces the risk of sensitive data being exposed in the cloud.

6.3 | Artificial Intelligence for Internet of Things Security

AI can also play a role in improving the security of IoT systems. AI-powered Intrusion Detection Systems (IDS) can monitor network traffic for signs of cyberattacks, such as unauthorized access or malware infections. These systems use ML models to detect anomalies in the data and alert administrators to potential threats.

In addition, AI can be used to secure communications between IoT devices and the cloud. AI-based encryption algorithms can dynamically adjust encryption levels based on the sensitivity of the data being transmitted, ensuring that sensitive data is protected without compromising performance.

7 | Energy Efficiency and Cost Optimization

Energy efficiency is a critical concern in IoT systems, particularly for devices that rely on battery power. AI algorithms, particularly DL models, are computationally intensive and can consume significant amounts of energy when deployed on IoT devices.

7.1 | Energy-Efficient Artificial Intelligence Algorithms

Researchers are developing energy-efficient AI algorithms that can operate on IoT devices with limited resources. Techniques such as model compression and quantization reduce the size and complexity of AI models, allowing them to run on low-power devices without sacrificing accuracy.

For example, in smart agriculture, sensors deployed in the field need to operate for extended periods without human intervention. Energy-efficient AI models can analyze data from these sensors and make decisions (e.g., when to irrigate crops) while minimizing battery consumption.

7.2 | Cloud-Based Cost Optimization

Cloud platforms offer various tools for optimizing the cost of processing IoT data. Serverless computing services, such as AWS Lambda and Azure Functions, allow IoT applications to scale dynamically based on demand. This dynamic resource allocation reduces costs by ensuring that cloud resources are only used when needed.

For example, a smart home system may only need to process data from motion sensors when someone is present in the home. By using serverless computing, the system can scale down when the house is empty, reducing costs without compromising functionality.

Cloud platforms also provide cost monitoring tools that allow businesses to track their usage and optimize their spending. These tools can automatically scale down services during periods of low demand, reducing operational costs and improving energy efficiency.

8 | Current Trends and Future Research Directions

The field of AI-enhanced IoT systems is rapidly evolving, with new technologies and trends emerging that promise to further revolutionize data processing and decision-making.

8.1 | AIoT: The Convergence of Artificial Intelligence and Internet of Things

The convergence of AI and IoT, known as AIoT, is one of the most significant trends in the industry. AIoT systems integrate AI into IoT devices and platforms, enabling autonomous decision-making and real-time

optimizations. These systems are capable of processing data locally on IoT devices, reducing the need for constant communication with the cloud and improving the system's overall efficiency.

AIoT is being used in a variety of industries, including healthcare, where AIoT devices monitor patient health in real time and make automatic adjustments to treatment plans. In manufacturing, AIoT systems optimize production lines by analyzing data from IoT sensors and making real-time adjustments to machinery.

8.2 | The Role of 5G in AIoT

The rollout of 5G networks is expected to have a significant impact on AIoT systems by providing high-speed, low-latency connections between IoT devices and the cloud. 5G networks will enable real-time AI processing for applications such as autonomous vehicles, where split-second decisions are required to ensure safety.

In smart cities, 5G will enable real-time monitoring and optimization of public services, such as traffic management and energy distribution. AIoT systems will be able to analyze data from thousands of sensors simultaneously, providing cities with the ability to respond to changing conditions in real-time.

8.3 | Future Research Directions

Several areas of research are likely to shape the future of AI-enhanced IoT systems:

- I. Scalability of AI models: As the number of IoT devices continues to grow, there is a need for AI models that can scale efficiently across large, distributed networks. Researchers are exploring new architectures, such as distributed AI and edge AI, that allow AI models to operate across multiple devices without relying on centralized cloud servers.
- II. Reducing AI model latency: Low-latency AI models are critical for time-sensitive IoT applications, such as autonomous driving and industrial automation. Techniques such as model distillation (where a large AI model is simplified into a smaller, faster version) are being developed to reduce the time it takes for AI models to make decisions.
- III. Ethical AI in IoT: As AI becomes more integrated into IoT systems, ethical concerns about AI decision-making are gaining attention. Researchers are investigating frameworks for ensuring that AI-enhanced IoT systems are transparent, fair, and accountable.

9 | Conclusion

The integration of AI into cloud-driven IoT systems has the potential to revolutionize how organizations process and analyze data. AI technologies such as ML, DL, and RL enable IoT systems to make intelligent, real-time decisions, improving efficiency, reducing costs, and enabling new applications in industries such as healthcare, manufacturing, and transportation.

However, challenges such as data security, privacy, and energy efficiency must be addressed to fully realize the potential of AI-enhanced IoT systems. Emerging trends such as AIoT, federated learning, and 5G networks promise to further enhance the capabilities of these systems, but ongoing research is needed to address scalability, latency, and ethical concerns.

As the field continues to evolve, AI-enhanced IoT systems will play an increasingly important role in driving innovation across industries, transforming how we live, work, and interact with the world around us.

Author Contributions

Conceptualization, methodology, and project administration by R. B. M.: software and validation by S. S. J.: investigation and writing—original draft preparation by A. S.: writing—review and editing by R. B. M. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data supporting the findings of this study are available upon request from the corresponding author. Due to the nature of the data, it cannot be publicly shared as it involves sensitive IoT datasets collected for research purposes.

Conflicts of Interest

The authors declare no conflict of interest. The funder played no role in the design of the study, in the collection, analysis, or interpretation of the data, in the writing of the manuscript, or in the decision to publish the results.

References

- [1] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>
- [2] Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. *Computer Networks*, 129, 459–471. <https://doi.org/10.1016/j.comnet.2017.06.013>
- [3] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE internet of things journal*, 3(5), 637–646. <https://doi.org/10.1109/JIOT.2016.2579198>
- [4] Vaquero, L. M., & Roderio-Merino, L. (2014). Finding your way in the fog: Towards a comprehensive definition of fog computing. *ACM Sigcomm Computer Communication Review*, 44(5), 27–32. <https://doi.org/10.1145/2677046.2677052>
- [5] Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial intelligence (AI) or intelligence augmentation (IA): what is the future? *Ai*, 1(2), 8. <https://doi.org/10.3390/ai1020008>
- [6] Fakheri, S. (2022). A comprehensive review of big data applications. *Big Data and Computing Visions*, 2(1), 9–17. <https://doi.org/10.22105/bdcv.2022.325256.1041>
- [7] Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2923–2960. <https://doi.org/10.1109/COMST.2018.2844341>
- [8] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., & Ivanov, V. (2019). Towards federated learning at scale: System design. *Proceedings of machine learning and systems*, 1, 374–388. https://proceedings.mlsys.org/paper_files/paper/2019/file/7b770da633baf74895be22a8807f1a8f-Paper.pdf
- [9] Neves, F., Souza, R., Sousa, J., Bonfim, M., & Garcia, V. (2023). Data privacy in the Internet of Things based on anonymization: A review. *Journal of Computer Security*, 31(3), 261–291. <https://doi.org/10.3233/JCS-210089>
- [10] Giordano, G., Palomba, F., & Ferrucci, F. (2022). On the use of artificial intelligence to deal with privacy in IoT systems: A systematic literature review. *Journal of Systems and Software*, 193, 111475. <https://doi.org/10.1016/j.jss.2022.111475>