

Enhancing AI-powered IoT systems: A lightweight edge-based framework for adaptive AI and interoperability

Name: Abdulaziz Saleh Mohammed Alhasil

Table of Contents

Abstract	1
1 Chapter 1: Introduction	3
1.1 Introduction	3
1.2 Background	5
1.3 Problem Statement	7
1.4 Research Objectives	8
2 Chapter 2: Literatures Review:	9
3 Chapter 3: Research Methodology	10
3.1 Architecture Design	10
3.2 Lightweight Model Development	10
3.3 Interoperability Mechanism	11
3.4 Expected Outcomes:	11
3.5 Gantt chart:	13
4 References:	14

Abstract

Combining the wide-ranging connectivity features of IoT with the sophisticated analytical power of artificial intelligence (AI), AI-powered Internet of Things (AIoT) systems propel the Fourth Industrial Revolution. Still, the widespread use of these technologies in important areas including healthcare, smart cities, and industry encounters major barriers that prevent them from realizing their whole potential. The performance-resource gap is the first challenge; here, sophisticated artificial intelligence algorithms—particularly black-box deep learning models—use enormous computational power disproportionate to the limited capabilities of peripheral devices, which are often battery-powered and include basic processors. The second difficulty is the lack of technological compatibility among heterogeneous systems, which causes isolated data islands to appear thereby slowing information transmission and network latency rise.

This study seeks to offer innovative solutions to these problems by creating a small AI-powered Internet of Things architecture based on edge computing. This approach uses an engineering approach meant to simplify difficult mathematical models and lower the quantity of calculations without sacrificing result correctness or security reliability. By building an intermediary layer that standardizes communication languages across several devices and platforms (such as MQTT and CoAP), this plan mostly seeks to improve the security of low-power networking protocols (RPL) while addressing interoperability problems.

Starting with a thorough analysis of prior research, the research methodology uses a quantitative experimental method and then creates a complex Edge-IIoTset dataset industrial IoT environment simulation. The suggested framework's capacity to balance power consumption and performance will be examined as will its effectiveness against node forgery and DDoS attacks.

The predicted scientific contribution of this study is to offer a practical model able of lowering power usage by at least 22% relative to conventional models and minimizing data latency to under 10 milliseconds, therefore guaranteeing immediate reaction in critical situations. Moreover, the study seeks to advance green AI by lowering the carbon footprint produced by extensive processing activities, therefore offering a sustainable and scalable way for using trustworthy, transparent, very efficient AI-powered Internet of Things systems in handling technical and environmental resources.

1 Chapter 1: Introduction

1.1 Introduction

With the emergence of the AI-powered Internet of Things, or the concept of the intelligence of things, over the past ten years, as a result of integrating the Internet of Things with the analytical capabilities of artificial intelligence (AIoT), this has led to a major change in global technology. The concept of the Internet of Things is no longer just a means of collecting and transmitting data, but thanks to its integration with artificial intelligence, it has become a system capable of making decisions, having future visions, and making predictions independently. This situation is changing important fields such as telemedicine, AI-driven manufacturing, and how smart cities are managed sustainably. This integration ultimately aims to create intelligent environments capable of fulfilling human needs with speed and precision, thereby improving and increasing operational efficiency and reducing material and human waste.

However, this rapid growth has faced technical problems and obstacles that threaten the viability of these systems and their widespread adoption. Most of these current systems are designed to operate in massive computing environments based on the principle of cloud computing. Cloud computing is characterized by virtually unlimited processing power, and this is where the main problem lies. This approach results in latency and slow system response times. Furthermore, attempting to use these cloud computing algorithms and apply them to peripheral devices—devices near the data source such as sensors and microcontrollers—leads to significant variations in resources and performance. Running complex AI models on these devices is a problem the field suffers from, as peripheral devices are inherently limited. Energy and memory consumption are significant challenges. Using these models leads to rapid battery drain and high processor strain, reducing lifespan and increasing maintenance costs.

Furthermore, technical compatibility presents another hurdle. The global market lacks standardized protocols for connecting numerous products and systems, resulting in monopolistic systems unable to easily exchange data with each other. This increases network response times and data loss in critical situations, particularly in security and medical applications.

This underscores the urgent need for this research, which aims to create a lightweight engineering framework that balances energy efficiency with high-performance security risk detection. This framework will ensure the stable and secure operation of diverse systems. This study not only seeks to improve technical performance but also to innovate a sustainable artificial intelligence concept that considers the physical constraints of the surrounding environment in its technological processes.

1.2 Background

The current technological revolution has led to a radical change in how digital systems interact with the physical world, a change embodied in the integration of what is known as Artificial Intelligence of Things (AIoT).

Previously, the Internet of Things (IoT) connected devices, collected raw data, and sent it to the cloud for processing. However, with the increasing volume of data and the growing need for real-time responses from applications, the need to integrate artificial intelligence directly at the edge of the network has become apparent. Thanks to this shift, systems possess cognitive capabilities that enable them to evaluate data locally and make complex, predictive decisions. This reduces their reliance on remote servers and accelerates system response times in environments requiring immediate processing, such as smart factories and medical applications.

The emergence of edge computing is a natural reaction to the difficulties edge devices face with traditional cloud computing, particularly regarding bandwidth, privacy, and response time. In applications such as industrial quality control, cybersecurity, or medical uses, any delay in data transmission is unacceptable and represents a significant obstacle in some systems. Therefore, the algorithm must be close to the sensor. However, achieving this close alignment presents a major engineering challenge, which most systems have adopted. Advanced AI models, often resembling black boxes, require massive graphics processing units (GPUs) and significant power resources, while the actual peripherals, such as ESP32 controllers or simple ARM processors, have limited memory and processing power and often rely on short-lived batteries.

Furthermore, digital sustainability is a key consideration in this study. Excessive energy consumption in data centers, computing, and powerful processors directly contributes to increased carbon emissions, contradicting the global trend toward Green AI.

The lack of interoperability in the design of IoT devices, stemming from the early stages of the IoT revolution, presents a significant challenge in building a stable and secure system. This is due to the existence of numerous different protocols and platforms, known as heterogeneous systems. The failure to achieve technical compatibility not only complicates management but also leads to the emergence of security vulnerabilities that can be exploited in network protocols such as RPL.

Therefore, the primary objective of this study is to highlight the necessity of achieving a balance between three important and vital elements: high performance, low power consumption, and integrated security. The transition from heavy to light models is not merely a technical decision to increase speed, but a strategic imperative to ensure that AI-powered IoT systems remain scalable, more professional, adaptable, and capable of operating in real-world environments while also being environmentally friendly.

1.3 Problem Statement

In the field of AI-integrated Internet of Things (AIoT), this study is a cornerstone for advanced systems in sectors such as industry, healthcare, and public services.

These systems contribute to increased efficiency and effectiveness in improving and accelerating decision-making through comprehensive advocacy power. However, these electronic devices face a fundamental and significant challenge: multi-performance practices and resource availability. In Industrial IoT (IIoT) environments, AI algorithms achieve up to 99.2% accuracy in object and location identification. However, this puts a strain on the processors and batteries of these edge devices, which are already resource-constrained.

This situation is exacerbated by a lack of standardized compatibility between devices, leading to system failures, sudden outages, and difficulties in data transfer between different platforms. This hinders the realization of a comprehensive internet vision in smart cities. These persistent outages lead to increased latency and data transfer times, a recurring problem in both the healthcare and industrial IoT sectors, requiring immediate and continuous action.

Existing solutions, as research has shown, include "edge computing" in agriculture, which relies on reduced storage and cloud processing. Another approach involves using artificial intelligence (AI) to minimize carbon emissions, as also discussed in research.

After examining all of this, a clear gap emerges: the lack of a lightweight AI framework. Therefore, there is a pressing need for a successful system or model that balances high security with low resource consumption, ensures compatibility across different device manufacturers, and guarantees the sustainability and efficiency of AI-powered IoT systems in complex real-world environments.

1.4 Research Objectives

Main Objective: This research aims to develop a robust AI-based technical framework specifically designed for edge computing. This framework aims to enhance the operational efficiency and security of industrial and service-oriented IoT systems, taking into account the limited resources, processing power, and energy consumption of edge computing-devices.

Specific Objectives:

To Evaluation: Identify weaknesses in traditional models when used in dynamic environments and conduct a comprehensive comparative analysis to assess the impact of modern AI algorithms (such as deep learning and reinforcement learning) on energy consumption and response times in resource-constrained IoT devices.

To Design: Create a lightweight AI model that uses adaptive learning or model compression to identify complex cyber threats and predict mechanical problems in real time with minimal strain on edge computing devices.

To Proposal: Develop and propose an intelligent communication protocol that facilitates interoperability, aiming to remove software barriers between different platforms. This will enable secure and seamless data exchange between devices manufactured by different companies within an AI-powered IoT network.

To verification: Using a working model or advanced simulation platforms for realistic scenarios in the healthcare and industrial production sectors, the effectiveness of the proposed framework was tested and verified, and its ability to balance device lifespan under challenging conditions with the accuracy of its results was assessed.

To recommendations: To ensure a secure and sustainable digital transformation, developers and decision-makers should be provided with a set of useful suggestions and technical guidance on how to integrate this framework into existing infrastructures.

2 Chapter 2: Literatures Review:

The Internet of Things (IoT) has changed big time lately. It used to be just about simple data monitoring. Now, it's all about smart IoT systems powered by AI. They build machine learning right into our virtual devices. Some recent studies suggest this shift is behind the latest Industrial Revolution. It's bumped industrial efficiency way up, like 99.8%. That means less wasted resources and fewer people needed.

But honestly, the "black box" thing is a major hurdle. It stops us from using it everywhere. That's especially true in sensitive fields like healthcare. People just don't trust AI conclusions if they can't understand how they got there. It makes them hesitate.

Resource constraints really come up when you look at the research. Those super complex algorithms They chew through an insane amount of processing power, That's what drains the battery and makes sending data to the cloud take forever. Cybersecurity work mostly went into R&D on encryption.

AI apps, like systems that spot intrusions, were a big deal too. Those efforts actually worked out really well. They managed to catch emerging threats 99.2% of the time. Honestly, analyses show AI's key for boosting those small smart energy grids in farming.

It might even cut the world's carbon footprint by 12%. That'd be something. We've gotta ditch these clunky, centralized communication systems. It's clear we need to move towards privacy-focused "light edge" models. They really cut down on how long it takes to get things done. Plus, they sort out those annoying compatibility headaches between different gadgets. Honestly, we probably need more digging into this stuff within a specific setup.

3 Chapter 3: Research Methodology

This study employs a quantitative experimental approach to design and test a lightweight AI framework for peripheral devices. The research will be conducted through a series of integrated methodological steps to ensure the objectives are met and results are accurately measured.

3.1 Architecture Design

The methodology begins with the design of a three-layer infrastructure:

Perception Layer: This layer includes sensors and IoT controllers that collect raw data.

Smart Edge Gateway Layer: This layer is the core of the research, as the proposed framework will be developed within it to minimize reliance on cloud computing and provide real-time processing.

Cloud Computing Layer: This layer is used exclusively for long-term storage and model retraining as needed.

3.2 Lightweight Model Development

To address resource constraints, knowledge distillation will be used. This process involves training a large "teacher" model on powerful servers to achieve high accuracy, then transferring this knowledge to a smaller, more efficient "student" model. Additional techniques will be employed, such as:

Model Pruning: To remove unnecessary neural connections and reduce the required memory size.

Quantization: To convert the Float32 model weights to Int8, reducing the computational effort on simpler processors.

3.3 Interoperability Mechanism

A semantic translation engine will be built within the Edge Gateway. This engine receives data packets from various protocols (such as MQTT from Tuya devices and CoAP from industrial devices), then converts them to a standard JSON format before passing them to the AI algorithm, ensuring a seamless, lossless flow of information.

3.4 Experimental Setup and Validation Environment

The proposed framework will be evaluated using a software simulation environment (such as Cooja Simulator or NS3) to model an RPL-based IoT network. The Edge-IIoTset dataset, containing real-world traffic and various cyberattacks, will be used. Efficiency will be measured based on specific criteria.

3.4 Expected Outcomes:

Several outcomes are expected from the development of the proposed AI and IoT framework. This framework is designed to address the performance gap between resources and interoperability constraints identified in the problem statement. Key technical and operational expectations include:

Improved Power Consumption:

Improved power consumption is a key expected outcome. Power consumption on peripheral devices will be significantly reduced using model trimming and quantization algorithms. The research aims to achieve a minimum 22% reduction in power consumption per inference compared to standard deep learning models. This will extend the battery life of remote sensors and industrial IoT nodes, contributing to overall sustainability.

Reduced Latency and Increased Response Speed:

This framework is expected to maintain local decision response times of less than 10 milliseconds by migrating decisions from cloud computing to the edge gateway. The improved system will provide immediate responses to security threats (such as DDoS or RPL attacks). This rapid response is critical for monitoring urgent healthcare and critical areas such as industrial automation, where every fraction of a second counts.

Achieving Technical Compatibility:

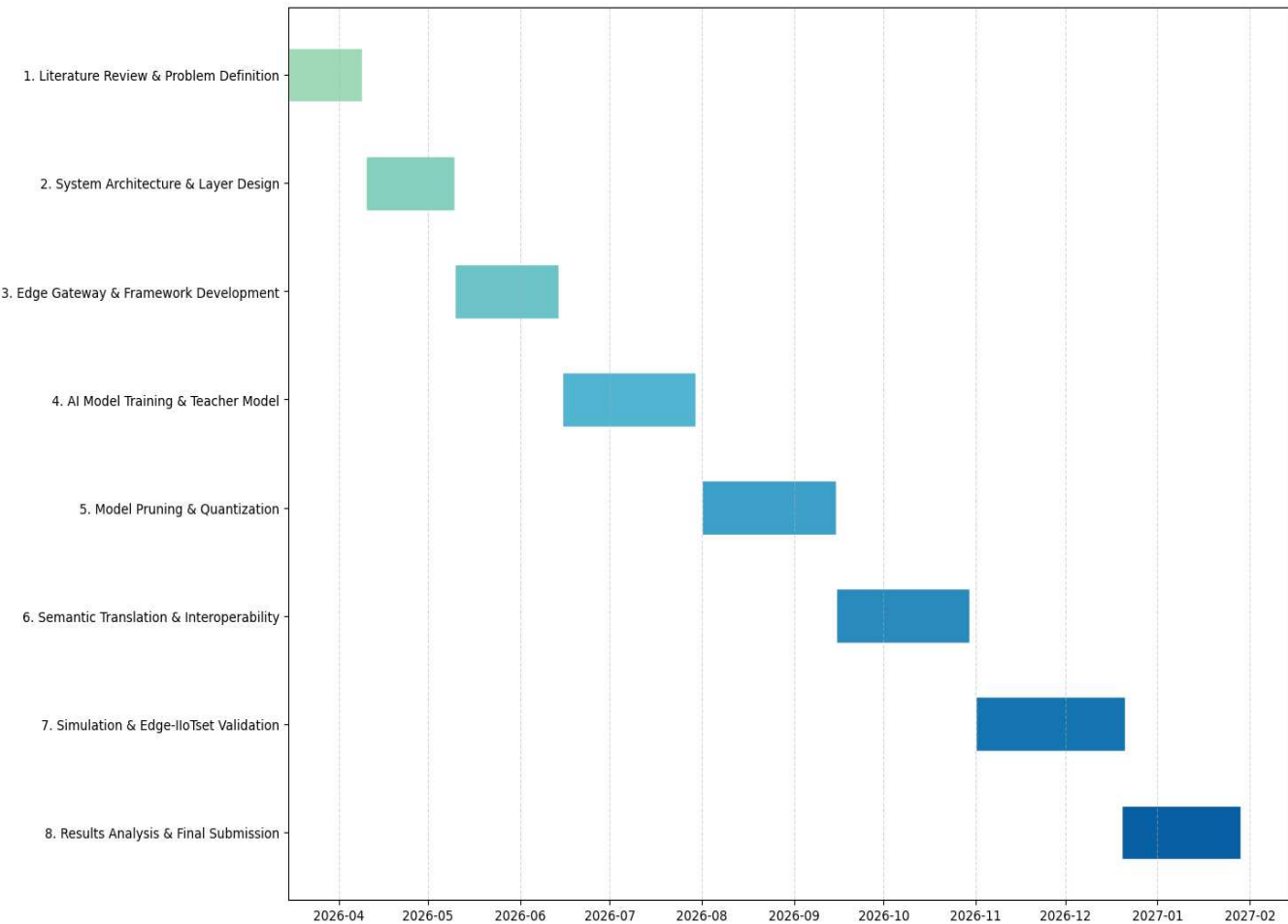
One of the most important expected outcomes is the successful integration of different devices, enabling them to exchange data within a unified framework. This enhanced semantic interconnect is expected to allow seamless data exchange between at least three protocols from different manufacturers (such as the Tuya-based MQTT protocol and the industry-standard Zigbee CoAP protocol). This compatibility will help eliminate manufacturer monopolies and provide a highly flexible and scalable infrastructure.

Reducing Environmental Impact:

In general, this framework will contribute to realizing the concept of green AI by reducing the carbon emissions associated with the transfer and processing of big data in cloud computing. A key benefit of this research is that it will provide a verifiable roadmap for reducing the overall energy consumption of all AI-powered IoT applications.

3.5 Gantt chart:

Research Roadmap



	Literature Review & Problem Definition	System Architecture & Layer Design	Edge Gateway & Framework Development	AI Model Training & Teacher Model	Model Pruning & Quantization	Semantic Translation & Interoperability	Simulation & Edge-IIoTset Validation	Results Analysis & Final Submission
Start	2026-03-15	2026-04-10	2026-05-10	2026-06-15	2026-08-01	2026-09-15	2026-11-01	2026-12-20
Days	25	30	35	45	45	45	50	40

4 References:

- * Al Rawajbeh, M., et al. (2025). Trustworthy Adaptive AI for Real-Time Intrusion Detection in Industrial IoT Security. Future Internet, 18(2), 82. <https://doi.org/10.3390/iot6030053>
- * Azmat, H. (2024). Artificial Intelligence and the Internet of Things: Transforming Connectivity and Smart Systems. ResearchGate Publication. https://www.researchgate.net/publication/398284167_Artificial_Intelligence_and_the_Internet_of_Things_Transforming_Connectivity_and_Smart_Systems
- * Hasan, M. M., et al. (2025). AI-Driven Quality Control in Manufacturing and Construction: Enhancing Precision and Reducing Human Error. Engineering Informatics, 31, 10270. <https://doi.org/10.25163/engineering.3110270>
- * Kasereka, S. K., et al. (2026). From IoT to AIoT: Evolving Agricultural Systems Through Intelligent Connectivity in Low-Income Countries. IoT Journal, 6(3), 53. <https://doi.org/10.3390/fi18020082>
- * Mustafa, A. (2025). AI-Powered IoT: Transforming Healthcare, Industry, and Smart Infrastructure. ResearchGate Publication. https://www.researchgate.net/publication/39828413_AI-Powered_IoT_Transforming_Healthcare_Industry_and_Smart_Infrastructure
- * Wang, Q., et al. (2024). Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence. Humanities and Social Sciences Communications, 11, 35205. <https://doi.org/10.1057/s41599-024-03520-5>