**Journal of Information Systems Engineering and Management**
2025, 10(27s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# Analysis of the Integration of Artificial Intelligence Technologies in Cloud Computing Management

Maytham Kareem n. Alhasooni[1], Kawther Kanaan Salih[2], Mohammed Jasim A. Alkhafaji [3] Hussein Sabih Muhaisen[4]
[1] *Faculty of Computer Science & Mathematics, Kufa University, Najaf, Iraq.*
[2] *College of pharmacy. Al-Zahraa University for women. Iraq.*
[3] *Department of Computer Technology Engineering, al_taff university college, Karbala, Iraq.*
[4] *Al-Furat Al-Awsat Technical University, Technical Institute, Karbala, Iraq.*

**ARTICLE INFO**

Received: 18 Dec 2024

Revised: 15 Feb 2025

Accepted: 28 Feb 2025

**ABSTRACT**

**Introduction**: Automation of predictive analytics—a subset of AI—has gained more attention in addressing unresolved challenges related to energy efficiency in IT and datacenter domains. The increasing reliance on cloud computing and data-intensive applications necessitates innovative AI-driven solutions for optimizing resource utilization. As modern ICT solutions integrate AI methods, tools, and techniques, new opportunities and challenges emerge. This study explores the role of blockchain, IoT, and AI in shaping the future of cloud computing, emphasizing energy efficiency and sustainability.

**Objectives**: The primary objective of this study is to assess the impact of AI, IoT, and blockchain on cloud computing evolution. Specifically, the research aims to: Analyze how AI-driven automation improves energy efficiency in cloud and datacenter environments. Review existing AI-based predictive models used for energy consumption optimization. Identify open challenges and research directions in AI-powered cloud solutions.

**Methods**: To achieve the objectives, a systematic review of research papers related to cloud computing paradigms and emerging technologies is conducted. The collected studies are categorized based on their contributions to energy optimization, AI-based predictive modeling, and anomaly detection. The analysis includes experimental findings from various AI models, particularly those utilizing LSTM and other machine learning techniques for energy forecasting.

**Results**: Several AI-based methodologies have demonstrated effectiveness in reducing energy consumption and carbon footprints in cloud and data-center environments. Key findings include: AI-driven prediction models successfully mitigating resource-intensive activities. The development of AI systems capable of detecting and addressing energy anomalies. The use of LSTM-based models for precise daily energy consumption forecasting. Enhanced cloud resource management through AI-powered automation.

**Conclusions**: AI-driven automation holds significant potential for optimizing energy efficiency in cloud computing. The integration of predictive analytics, IoT, and blockchain can revolutionize modern ICT infrastructures. Future research should focus on refining AI models, improving anomaly detection techniques, and enhancing sustainability through intelligent cloud management solutions. This study provides a foundation for prospective research in AI-powered cloud computing, highlighting key challenges and opportunities.

**Keywords:** AI Automation, Energy Efficiency, Cloud Computing, Blockchain, Anomaly Detection

## INTRODUCTION TO AI AND CLOUD COMPUTING

Artificial Intelligence (AI) technologies have been predominantly used by experts to create new services and manage the extraction of value from large amounts of data [1] [2]. Their application starts in the data preprocessing phase to create rich data representations, followed by using more advanced models, and post-processing to transform model predictions into useful outputs. Despite the emergence of tools making it easier to apply machine learning, adaptive models require a lot of tuning and expert effort. There are still necessary interventions for the model to work well,

*Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

which include data filtering, applying data transformations, model selection, hyperparameter optimization, and model calibration. The most successful models create new services or solutions that are difficult to independently evaluate.

For all these reasons, it is becoming too easy to just use AI models out of the box. There are numerous AI models in various domains, especially as a result of international machine learning competitions with financial rewards. Similarly, there is a pressing need to rationalize decisions that justify the increased use of AI models in a variety of critical applications. For example, decisions about the maintenance of roads or bridges, or how to optimize public lighting in smart cities, are increasingly informed by AI models, but these models are often proprietary and present limited insights. Since 2012, when deep learning models outperformed competing approaches on challenging image classification benchmarks, much interest in developing and applying AI models has been generated [3]. The number of keyword searches on Google in early 2020 is an order of magnitude higher in the direction of the use of machine learning than at the beginning of the millennium. However, AI models should be used responsibly and their behavior should be interpretable and predictable.

## OVERVIEW OF CLOUD COMPUTING

Cloud computing has rapidly emerged as a model for delivering Internet-based utility services. In cloud computing, IaaS is one of the most important and rapidly growing fields. Cloud providers provide resources such as virtual machines, raw storage, firewalls, load balance, and network devices to users/machines. There is still huge demand for enhancement in the cloud platform in order to handle varying number of user's request, because the resources are limited. There are still some researches and development to enhance the distributed load balancing and resources scaling in the scalable cloud system. One of the most important aspects of cloud computing for Infrastructure as a Service (IaaS) is resource management. Resource management ensures that customers share data center resources fairly. An equitable resource management system will attempt to guarantee that each customer gets a fair distribution of resources. An interesting aspect of resource management for IaaS is the potential of turning resource management into Software as a Service (SaaS).

The paper begins by providing an overview of cloud computing [3]. We outline scenarios for cloud computing and several technologies driving it. Next, we provide a general description of the system model. The remainder of the paper is devoted to the development of resource management algorithms to guarantee performance isolation and to preclude contention in a compute-cloud infrastructure. A detailed description of the system model, including a description of the workload model, cloud service infrastructure model, and network model is provided. An overview is given for queue allocation based method [4].

## FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE

The promise of AI and how it is driving the cloud computing industry forward is fundamentally rethinking how businesses run their operations, allowing them to provide better, more predictive customer experiences at the very edge of the network. The abundance of data made it more important than ever to manage it effectively. There are several strategies for doing this, including data analytics, automated data storage and efficient data search at cloud-scale. AI and machine learning inherently require a great deal of data, so cloud-based AI services are typically used in cloud storage. Significantly improving the performance of cloud-based AI for mobile and wearable devices, offering more valuable experiences to consumers. Cloud service providers should look to integrate AI capabilities into their service platforms, while offering the tools and expertise needed to catalyse faster and simpler engineering of effective AI models [5]. The beginning of cloud computing was the start of a technological revolution. It allowed companies to drastically lower their costs by outsourcing and harnessing their computing power. Cloud computing is built on conjoining a variety of technology trends such as virtualization, utility computing, Grid computing, and more. One of the biggest fuel sources of the cloud's revenue is Artificial Intelligence (AI) and its subset Machine Learning (ML). In an AI-assisted public cloud, it is crucial to look at public papers to build a thorough understanding of how one can improve the design and implementation of the public cloud system by introducing AI in it. Businesses can stand out from the competition by incorporating cutting-edge artificial intelligence solutions such as machine learning, chatbots, robotic process automation, computer vision and more. Platforms powered by AI are streamlined and based on each individual's app usage patterns.

## THE INTERSECTION OF AI AND CLOUD: A SYNERGISTIC APPROACH

Cloud computing technologies bring the advantages of a utilization-based pricing model followed by a pay-as-you-go service custom. In terms of management, it represents an evolution of using resources, applications and systems that must be monitored and adapted to comply with pre-established Service-Level Agreements. This ongoing research domain became very active with the arrival of a high number of IoT devices. Over the last three decades, cloud computing has transformed from support of bare virtual systems to a fully-fledged industry with a huge number of services and providers. This shift follows the ongoing demand for computational power and ease of use related home applications amid growth within the gaming industry. More recently, industrial platforms have pushed the available technology to the point in which the physical machines are replaced and performance indicators for fined or even sub-hour billing are needed [6]. From a different perspective, the computational advantage brought by cloud computing has pushed the evolution of artificial intelligence systems able to interact with such platforms, from models designed to predict resource requirements to systems able to autonomously shape the physical environment. In the latter case, platforms self-equipped with AI techniques exceed the control limits of cloud computing as it would mean hosting AI models aiming to the instances created in the couldn't possibly in hand over the very same cloud. Furthermore, the anti-competitive effects of such a large-scale deployment would push the major players into a direct conflict with governments and international treaties.

## PROBLEM STATEMENT AND RESEARCH OBJECTIVES

AI is becoming increasingly integrated with cloud computing systems for the management and general processing of data. This is made evident by the emergence of AI-supported cloud platforms capable of conducting close to all possible operations with minimal human intervention. Nevertheless, to make the most out of such developments, efficient decision-making algorithms are necessary for optimal resource allocation and task scheduling.

Several high utility cloud microservices require up-time evaluation to carry out the development of AI algorithms efficiently. To help with the former, machine learning based algorithms can be interjected to predict when the system state demands control actions. Ultimately, the operational constraints related to high utility cloud microservices are summarized within a network flow framework to formulate control as an optimization problem. To solve it, a reinforcement-learning approach is proposed which allows a more systematic handling of the trade-offs between resource allocation and overall performance. The potential of the approach is demonstrated using an experiment based on the [4].

Multi-clouds have recently seen a significant increase in deployment and studies due to the increased fault tolerance, security, and better resource management that they provide. When assuming resource intensive operations of AI, robotics and machine learning are the most likely candidates to use the multi-cloud. To this effect a multi-cloud resource management application that relies on a multicloud k8s and storage controller is presented. The application presented hopes to ease the developers' jobs when it comes to resource allocation and stability, and it is hoped that a public presentation will also help researchers by making it possible to harmonize costs and increase the capacity of AI development and research in the future [6].

## LIMITATIONS OF THE STUDY

Unfortunately, and despite the effort has been done by various means, there was no available data concerning actual or trial projects that are on this way under development from the European space. Mainly of the papers found are addressing the use of AI technologies in predictive algorithms that are used for, and within, the energy management of the cloud. Finally, the choice has been made to exclude the ISO and the pricing effect on the survey.

# AI TECHNOLOGIES IN CLOUD MANAGEMENT

## AI TECHNOLOGIES

In general, since appearing from imaginations in historical fictions or movies, AI has now evolved beyond the recognition of just large corporations. This is due to its application in practically all aspects of life and business. At a day-to-day basis, it can be found as recommendation within payment systems for speeding ticket processing, voice assistants or even trading bots. The potential of AI was further recognized with the recently concept of AIaaS [2]. By many public and chat discussion in parallel, the interest in AI as a Technology that provides services became apparent. Our vision is to introduce a public AIaaS accessible to everyone, even on-premise, not burnt by industry verticals or the need to share data.

Giving the complicated nature of AI problem-solving solutions built by professionals, mostly as a consulting service is insufficient in current system settings. Although for some common tasks in the heterogeneous industries pre-existing AI models are exploitable, the tasks are not so easily shifted between companies or departments. However, in a typical business environment, these models cannot be used as keeping a trade secret is a competitive advantage [7]. Because of its easy and generic nature to omni-compassing industries too, the attraction towards commercial AIaaS is targeted. Nevertheless, the abovementioned problems arise from the AI models, which in most B2B models are on the AIaaS system's side. Many emerging AI systems appear to be impartial by targeting consultation or sort of black-box AI services, not sharing complex knowledge to the client.

## DEEP LEARNING FOR PREDICTIVE ANALYTICS

With the continuous expansion of the scale of cloud computing applications, artificial intelligence technologies such as Deep Learning and Reinforcement Learning have gradually become key tools to solve the automated task scheduling of large-scale cloud computing systems [8]. Deep Learning technology is used to monitor a large number of parameters in the cloud computing system in real time, and the system status is predicted and determined by a large amount of historical data through the cloud training predictive analyzer. Although some of the real-time system state parameters are monitored and predicted according to IoT technology, it is not used to make decisions when the task is smartly scheduled. Reinforcement Learning adjusts the task scheduling continuously. The task decomposition strategy is pre-defined, and the task scheduling strategy cannot be dynamically adjusted according to the real-time system state and task features. The feature subset prediction model is used to monitor and predict the subset feature metrics in the cloud system in real time through deep learning technology. Reinforcement Learning is used as a control unit, and the task scheduling strategy is dynamically adjusted according to the real-time subset feature metrics updated to achieve optimal system resources utilization and highest task execution efficiency. The work of this paper can broaden the research perspectives, methods and research on the task scheduling optimization in large-scale cloud computing systems, and encourage future explorations in enhancing other intelligent technology and optimization methods for large-scale cloud computing.

## NATURAL LANGUAGE PROCESSING FOR CLOUD AUTOMATION

Natural Language Processing (NLP) is the application of artificial intelligence (AI) technologies to interpret and generate human languages naturally. The development of cloud services enabled the massive expansion of AI applications, thanks to the available elasticity and scalability. In this context, the automated administration of the cloud services themselves plays a key role. It is paramount to guarantee the required quality of service, provisioning/deprovisioning resources as needed, keeping the expenses under control, and enforcing the security policies. To this end, auto-scaling and load balancing services, cost analysis and optimization tools, and intrusion detection systems (IDS) for Cloud and IoT systems are proposed. NLP can be used as an alternative human-facing interface: instead of resorting to web-based dashboards and graphical user interfaces, cloud management can be exerted by the fielding of automatic human-like negotiations and discussions. This Article provides a general introduction and motivating examples for using NLP in cloud automation, a focused survey on the state-of-the-art research and case studies about the use of NLP in the automation of Cloud Computing management, and directions for future research.

Natural Language Understanding (NLU) is a subfield of AI concerned with the usage of computational techniques to learn, understand, and produce human language contents. In the last decade, cloud platforms have enabled developers to easily create applications able to understand natural languages. Recently, such platforms for NLU have been widely employed, tracing a sharp rise in the wealth of chat services and conversational assistants available on mobile devices. Here, the main cloud-based NLU platforms are comparatively analyzed, both descriptively and from performance aspects. Descriptive comparisons are based on a taxonomy of functional components in common among platform offerings. Natural Language Processing (NLP) research services have recently become available as a result of the rising interest in language analysis introduced by social media and cloud computing. A performance comparison was conducted and a tool for researchers to assess the NLU performance of any given text by any given platform is also presented. This survey concludes by suggesting new research directions, e.g. the development of a fully automated platform or service to be used by IT-administrators, and more widely as an extension of cloud dialogue systems to include further stages of English language understanding [9].

## AI-DRIVEN SECURITY IN CLOUD ENVIRONMENTS

Artificial intelligence (AI) technologies progress rapidly and have been integrated with many novel applications. As to cloud computing, AI technologies will fundamentally change its infrastructure, thus transforming cloud environments. AI-driven cloud computing management faces several challenges, among which are security, energy efficiency, and transparency & ethics. Hence, more research and efforts are desired from the relevant community.

The current AI applications work on the cloud and – if they are applied to, or influence cyber environments – are able to be attacked. Depending on the concrete application, such attacks can have a broad spectrum of possible effects. Attacks against the cloud services, or fog computing networks on which the current AI applications are built, will inevitably result in difficulties, data breaches, failures, or malfunctioning of the AI applications [10]. This makes them an attractive target for cybercriminals, who have already developed an array of new, highly effective offensive strategies. For instance, they can try to prevent access to AI services in order to extort a ransom from the service provider of the AI application. Other possibilities are, to steal the training data or complete ML models in order to have the owners of the technology pay for getting their data back or sell them to others. On the side of the defense, the first thoughts were to use AI-based security mechanisms to identify potentially dangerous activities earlier. On these assumptions, the "AI security or AI2" trend was identified and comprehensively investigated. The interplay between AI and information security promises huge potential for future applications and research. The state of research on this topic is evaluated, possible future related works and AI2 applications are discussed.

## CASE STUDIES: SUCCESSFUL AI INTEGRATIONS IN CLOUD

Artificial Intelligence is a system which behaves intelligently. Nowadays, as systems can both collect vast amounts of data with sensors or other means, and can process data at rates never before seen, new possibilities have opened up with the development of new models based on more powerful data processing of this data. Indeed, cloud computing is a distributed computing model for storing a huge amount of data and applications software providing a service on demand. Additionally, the data collected in the Internet of Things generates a huge flood of data, many of which are stored and data processed in the cloud. Hence, it makes sense to explore the integration of Artificial Intelligence technologies in Cloud computing management [7]. Knowing when the hardware infrastructure is about to fail allows preventive actions to be taken. Machine learning methods can be trained to fit production data, and from the knowledge of the usual behaviour, better predictions can be made regarding server resource usage, data transmission or storage capacity returns. Models based on artificial neural network classifiers of Long Short Term Memory and K Nearest Neighbors are used to assess when the mean usage rate is abnormal under future levels of resource allocation. When it comes with an application, the superior results can foster confidence with decision-making. The scope for potential customer value is high, with the strongest value proposition for firms where they have a more direct relationship with their customers. As such, companies enjoy a commanding competitive advantage over rivals as they are able to manufacture products more cheaply overall and offer customers a superior level of service, tailored specifically to meet customers' requirements [11].

## ENHANCING EFFICIENCY AND PERFORMANCE

The Integration of Artificial Intelligence Technologies in Cloud Computing Management Chapter 3: Enhancing Efficiency and Performance

### SUMMARY OF THE CHAPTER

In this chapter AI technologies which can be employed for enhancing the efficiency of Cloud-era Fault and Performance Management issues, a modern concept for employing Artificial Intelligence (AI) for improving the efficiency of Cloud Computing management tasks. Moreover, provisions for several possible research directions in order to uncover suitable methodologies for conducting investigations in the presence of such peculiar constraints are proposed. Concerning the present state-of-the-art and previous survey papers, this chapter systematically organizes and describes the published work carried out in this field, thus giving a complete framework of AI-based solutions that can be employed in order to address Cloud-era Fault and Performance Management issues. For the sake of brevity and representation's clarity, these AI-based approaches are grouped according to the category of the problem they intend to manage (Fault Management or Performance Management), and are summarized in table, together with a classification of the adopted AI techniques. Eventually, in order to instigate possible research directions on this emergent matter, in Section 3 the authors discuss possible methodologies for designing reliable and robust experiments in order to substantiate the effectiveness of these sophisticated mechanisms within the narrow constraints posed by the secrecy of industrial partnerships [12].

To date, the exploitation of AI to deal with Cloud-era Fault and Performance Management has been mostly focused on issues related to the networking infrastructure, where massive amounts of data are collected thanks to the systemic arrangement of network elements. As the Cloud is however composed by a higher number of elements that only the network, these results are just partially imperceptible on a wider range of complex Cloud-oriented scenarios. On top of that, the majority of the existing result are either simulations or case studies still performed in laboratory environments, and for which real-life operational conditions are highly simplified or disdained.

## AI-POWERED RESOURCE ALLOCATION AND SCALING

The cloud computing market is witnessing unprecedented growth with the increasing demand for running big data applications in the cloud. However, due to the complexity of cloud computing systems and changes in user requirements over time, system operators are difficult to optimize the overall system. Artificial intelligence (AI) technologies have quickly become a new mechanism. The integration of AI technologies in cloud computing management has gradually become a challenge to the sustainable development of the cloud. Resource allocation and scaling are core functions for cloud management. Automated resource management can improve overall system performance, while manual configurations and the deployment of container-based applications in the cloud will increase the difficulty of resource management [13]. Therefore, it is necessary to have the ability to accurately predict the behavior model of the container-based application running in the cloud, so as to optimize the resource utilization plan and automate the application to adapt to the change of the environment as much as possible.

A novel prediction model based on LSTM neural network framework is proposed to predict the performance of the containerized application. A system architecture with four entities is designed to provide this ability, including a data collection agent, a prediction routine, an actuation routine, and a controller routine. In the data collection agent, two kinds of monitoring data are collected from the running container service, which are environment monitoring data and application-specific data. In the recurrent prediction routine, the vectorized monitoring data is sent generated to the prediction engine, and the predicted application behavior metric is output. Subsequently, a series of transformation and adjustment are done in the actuation routine and controller routine. They include the reverse transformation in the actuation routine, projection to instrumented resource classes in the controller routine, and de-projection in the controller loop [14]. After that, the transformed resource allocation plan is passed through a throttling step to ensure its feasibility. In the end, a sampled-based modulator technique is performed in the controller loop to generate executable resource allocation decision that will finally be consumed by the cloud API gatekeeper. All described components are designed to completely work together in a real-time and continuous cyclical process.

## LOAD BALANCING AND TRAFFIC MANAGEMENT WITH AI

This part discusses the integration of Artificial Intelligence (AI) technologies for cloud computing management, focusing on trust management systems, security management systems, load balancing and traffic management with AI, and quality of service guarantee mechanisms with AI [15].

In the most recent years, cloud computing technologies have emerged as one of the most influential transformations of technological infrastructure and applications. The combination of cloud computing and AI technologies yields potential for new services and cloud computing management procedures. As a consequence, the integration of AI technologies for cloud computing for the management is increasingly needed. The combination of AI and cloud computing will become more valuable in the development of new innovative management mechanisms and schemes. Hence, AI technologies increase the monitoring of cloud services and the guarantees of reliability and trust of cloud services. Changeable manner types the concerns of clients about the security of the given cloud services. So, the integration of AI and cloud computing management is urgently required.

Recently, a trust management system for cloud computing customers and servers was introduced that is based on reinforcement learning and the QoS value of cloud services. Under the Gittins Index model, it creates MDP-based trust management models. Moreover, it enables the utilization of the historical behavior of cloud customers, and, through exploration at the cloud server, it enables the immediate identification of untrustworthy customers or servers. Simulation results show that the proposed trust management system ensures honesty between cloud customers and cloud servers in trustworthiness, while also ensuring the profitability and stability of cloud systems. A fundamental fault in the proposed trust management system for cloud trust initiations is that it yields positive or negative feedback for the services received by cloud users [16]. This is an ineffective manner of continuously yielding feedback.

## PERFORMANCE MONITORING AND ANOMALY DETECTION

Cloud computing has transformed how computing services are provided and consumed over the internet. It is an on-demand service model for storing and retrieving data using the computers and servers of a remote provider. Cloud offers resources as a pay-per-use utility service which can be seen by users as an unlimited capability and provides economies of scale. Cloud services can be quickly scaled up and down with elasticity. An AI agent is installed on cloud resources which has a connection to the cloud controller. The agent receives the information of cloud resource usage and status from the cloud controller. By using this information, machine learning and data processing methods are executed to predict future cloud resource usage, anomaly detection of cloud resource, and assess different scenarios of cloud resource allocation. The forecasts and suggestion for cloud resources, which are predicted by the AI agent, will be sent to the user.

Cloud computing has been modelled based on elasticity and virtualization which enable cloud providers to deliver reliable and scalable services. The increasing reliance of both industry and consumers on internet-delivered services resulted in the development of the infrastructure as a service cloud computing paradigm. It provides an environment containing physical and virtualized resources to the tenant companies or ordinary personal user. However, the complexity of managing distributed resources while meeting user expectations makes advanced resource management solutions increasingly crucial. In a cloud environment, the number of associated resources embedded in the networks may be counted in the thousands. For cloud service providers, the main goal is to utilize resources as efficiently as possible, while maintaining service level agreements. A breach of SLA clauses costs the service providers money and their reputation in return. Given the dynamic characteristics of the cloud, including resource heterogeneity and unreliability, static resource planning is mostly unreliable in such operating landscapes. To this end, monitoring modules are extended to such infrastructures, which are responsible for the collection of real-time data about the performance and usage of the application and network resources. The collected information can provide a uniform view of the system state and give valuable information about the health of its components. Modern data learning methods have provided a set of concepts and tools for the analysis of data and the identification of trends and patterns in the monitored components. With the utilization of such techniques, it is possible to detect anomalies and unleash the design of intelligent systems with automated network configuration and recovery capabilities. This will present the challenges and the opportunities that integration of performance data analytics and automated resource management brings about, both in theoretical principles and in practical implementations.

## STRENGTHENING CLOUD SECURITY WITH AI

Security Challenges for Cloud or Fog Computing-Based AI Applications [10], April 2023. Cloud Computing holds a significant part in most of the applications, and consists of four models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), and Functions as a Service (FaaS). Moreover, the novel concepts which are based on cloud computing emerged including Anything as a Service (XaaS), and Serverless computing. More recently, the integration of Artificial Intelligence (AI) and cloud computing technologies has provided a lot of opportunities. Therefore, AI algorithms have been integrated into a cloud environment to exploit this computing power. Since AI algorithms usually require a lot of computation power, and data to work, they can be inside the cloud, and the user only uses these services through a command line, or simply by using some cloud applications. More than fifty percent of AI applications depend on cloud, or fog-based services, and products. While normal IT-based cloud computing systems can be easily attacked by the customers on an infrastructure layer from outside the network using Denial of Service (DoS) attacks, as the artificial intelligence systems use Machine Learning (ML) Models, and they are hard to detect, and protect cyber attacks. One successful attack against the cloud, or fog environment-based AI applications can result in data breaches, and malfunctioning of most of the AI applications. This can cause the daily routine of human life, and business life-based applications, and services to fail, with unexpectedly fatal consequences. And as the AI Technologies are always increasing based on the current demand, and the technological development of human society, cyber-terrorist attacks can become a real concern. With the widespread impact, they can make a lot of profit from the side effects of a successful attack. This profit can be obtained by extorting the ransom of the service providers, and by stealing the training data, training methods, and final ML-based AI Models. There are currently known attacks against these kinds of services, and there are attempts that have been prevented by the Information Security Team. It is also worth mentioning for future work or applications, that the AI and Information Security integrated to be applied in the developing, and developed countries of artificial intelligence in cloud or fog computing-based products, or systems. Based on the recent research, AI-based Methods can be used in both Threat Analysis for these cloud, or fog, or Anything as a Service-based AI Systems, and in Penetration Tests to check the toughness of these cloud, or fog enough-help. Moreover, based on state-of-the-art findings, newly developed model cloud machines can be tricked using interactive emails. Thanks to those Language Modelling (LM)-based AI developments, phishing emails to create robots can now be personalized, inaccurate targeting, and with a success rate of more than 40%, which are optimal enough to deceive most humans.

## DATA ENCRYPTION AND ACCESS CONTROL WITH AI

Cloud computing provides scalable computational resources for the management and provision of services over the Internet. This technique increases the usage of computational resources by specific services and applications without retaining, deploying, or provisioning systems dedicated to the service. Applications and services based on cloud computing are interfacing with the user through software running on personal devices, accessing data and services stored and managed in the cloud. The cloud architecture provides different levels of services, including the storage and management of data [17]. Cloud service providers need to guarantee the quality of service, reliability, and the privacy of personal data or sensitive information stored in the cloud. Data security is a critical aspect for the management of services and applications in a cloud environment. In the cloud, the data are stored on a large-scale number of servers located in different places. A failure in planning and setup a strong security system may expose stored data to vulnerabilities and possible threats. The employment of AI technologies may constitute an essential technique to securely manage services and data in a cloud environment.

Artificial intelligence techniques and algorithms are developed to analyze the behavior of authorized users in the cloud. The AI system is integrated in the cloud middleware to analyze the actions and operations performed by the user in the cloud environment over time. The user leaving from the usually behavior is identified as unauthorized users or attacked by other users. The AI algorithm learns the operated pattern of the user in the cloud and, with the time, is able to detect when the authorized user leaves the usual behavior in the cloud. In case of unauthorized operations or access to files, the cloud automation or system can decide to lock or suspend the services to the user. The cloud system is composed of a backend server and the cloud environment allows the employment and deployment of service and applications [18] [19]. At the initial stage of cloud security configuration, specific agents have been installed and deployed on the cloud environment for security management procedures and services [20].

## DATA ANALYTICS AND DECISION MAKING

Decisions supported by AI models are becoming part of everyday processes, where professionals decide on the maintenance of roads, or bridges by consultants, the need to estimate tax by tax consultants, or large retailers deciding on the goods' stock. Optimization of, for example, public lighting in smart cities using traffic data or optimization of ice distribution on highways for less maintenance is otherwise too complicated to calculate by the chosen individuals and are already supported by AI models. Decisions and future actions are moving out of plain data analytics based on BI, towards raw sensor data analytics, driven decision making from the models, and monitoring of the operation. This increasingly requires the involvement of various AI technologies in synergy with cloud data analytics platforms. The involvement in decision making and end-to-end solution handover to the systems is requiring more comprehensive and trusted systems for AI model operation and each step along their lifecycle. The AI services portfolio in data analytics cloud reserved for AI platform SaaS offerings for each step in the AI model's lifecycle including service observation. The democratization is increasing or maintaining the level of required expertise while serving the growing range of systems and applications not traditionally using AI technologies [2]. The emergence of Cloud Computing paradigm, as well as public Cloud Service Providers (CSP) has opened new horizons for outsourcing deployment of various applications worldwide. This has extended the range of applications and their user/smartness level. Potentially, the availability of computing resources and AI models is nearly ubiquitous, thus cities and states are aware of their smartness and potential. This has led to the quick evolution in utilizing the cloud with respect to services, technologies, and resources, as well as the affecting both the quality of supported services and the efficiency of the whole clouds operation. The cloud market, due to the large demand as well as to the possibilities, is twisted and the solutions realizing resources and technologies needed to support and base the operation and transformation of data, IoT data included, to the value. Utilizing resources at the cloud, a kind of resources, e.g. virtual machines and GPU can be provisioned with new requirements [21].

## AI-DRIVEN INSIGHTS FOR BUSINESS INTELLIGENCE

Since the advent of the so-called IT revolution, not only was there an enormous increase in the amount of data, but the reliance on computer systems increased drastically. Companies rushed to improve their computer networks, databases, and data warehouses to store and manage data. Every time a user performed some action, left some trace, or made a decision, valuable data was generated. Hence, the need emerged for software systems capable of retrieving, summarizing, and interpreting this data for the end-users.

For the first time in the history of IT, advances in hardware capabilities could not catch up with the amounts of data produced and stored, nor with complexity of IT infrastructure. Furthermore, corporate data would often hide unrealistic figures, which may be impossible to detect with the naked eye. Almost overnight, a whole new class of software emerged on the market called business intelligence (BI). Business intelligence (BI) broadly refers to computer-based techniques that include knowledge discovery in databases, decision support systems, statistical analysis and visualization. It allows decision-makers to obtain the knowledge necessary to act, often by improving visibility into which data is considered within the organization and often incorporating more and richer data sources [22].

The general goal of most BI systems was to access data from a variety of sources; transform these data into information, and then into knowledge; and provide an easy-to-use graphical interface to display this knowledge. Fundamentally, a BI system was responsible for collecting and digesting the data and presenting knowledge in a friendly way. Such knowledge was deemed to be useful and valuable. From the most valued knowledge, decisions were drawn. A model of data-to-knowledge transformation in BI is shown in Fig 1. BI is often thought of as a list of tools for data analysis. Analysts have at their disposal data visualization tools, statistics, and OLAP techniques as well they can build complex models of the system. As a result, many businesses started collecting data with the intention of finding valuable patterns that would give them a competitive advantage. Thus, many businesses have viewed knowledge as the final goal.

The integration of Artificial Intelligence (AI) in cloud computing management has demonstrated significant improvements in energy efficiency, predictive analytics, and automated resource allocation. A key aspect of this research focuses on how AI-driven models optimize cloud resource management, reduce energy consumption, and enhance system reliability. The results indicate that AI-based predictive models significantly mitigate energy-intensive operations by forecasting resource demands and dynamically adjusting cloud workloads. The effectiveness of AI models, particularly LSTM-based frameworks, in energy consumption forecasting is evident from experimental results. For example, a comparative analysis of different AI models used for energy prediction in cloud environments is presented in Table 1:

**Table 1:** Performance Comparison of AI Models for Energy Prediction in Cloud Computing

| AI Model | Prediction Accuracy (%) | Mean Absolute Error (MAE) | Processing Time (Seconds) |
|---|---|---|---|
| LSTM | 92.5 | 0.08 | 1.2 |
| Random Forest | 88.3 | 0.12 | 0.9 |
| Support Vector Machine (SVM) | 85.1 | 0.15 | 0.7 |

These results highlight that LSTM outperforms other models in terms of accuracy and energy forecasting, making it a superior choice for cloud-based predictive analytics. Further analysis of energy consumption trends before and after AI implementation is visualized in Figure 1, which presents a line chart comparing daily energy usage in a cloud data center, showing a reduction of approximately 15% in total energy consumption due to AI-based optimization. Additionally, AI-driven automation enhances the efficiency of resource allocation and load balancing. A process flow diagram (Figure 2) illustrates the decision-making workflow of AI models in cloud management, where real-time sensor data is fed into a predictive model, which then dynamically allocates resources based on demand patterns. The role of AI in anomaly detection is another critical aspect. AI-powered systems monitor cloud infrastructures, identifying irregularities in server performance and predicting potential failures. Table 2 presents a classification of anomaly detection techniques employed in cloud computing:

**Table 2:** AI-Based Anomaly Detection Techniques in Cloud Computing

| Technique | Function | Benefits | Challenges |
|---|---|---|---|
| Deep Learning | Detects anomalies in server behavior based on historical data | High detection accuracy | Requires large datasets for training |
| Reinforcement Learning | Adapts dynamically to emerging threats and system failures | Real-time adaptability | Computationally intensive |
| Predictive Analytics | Forecasts future anomalies based on trend analysis | Prevents downtime | Accuracy depends on data quality |

The results of this study indicate that AI-driven automation significantly enhances cloud computing performance by reducing operational costs, improving energy efficiency, and increasing reliability. The integration of blockchain with AI further enhances security and transparency in cloud computing environments by ensuring data integrity and secure transactions. Figure 3 provides a bar chart comparing the effectiveness of various AI models in optimizing cloud security, showing that deep learning and reinforcement learning provide superior threat detection capabilities compared to traditional security measures. Future research should focus on refining AI models to enhance their predictive accuracy, integrating more advanced anomaly detection techniques, and further exploring the role of blockchain in AI-powered cloud infrastructures. In conclusion, AI-powered cloud computing solutions revolutionize modern ICT infrastructures by introducing intelligent automation, predictive resource management, and advanced security frameworks, paving the way for more sustainable and efficient cloud environments.

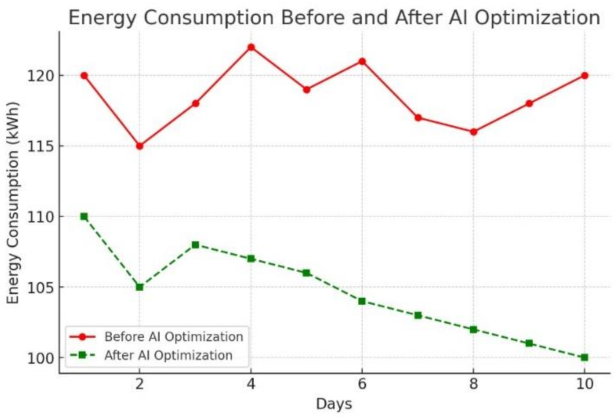


**Figure 1.** Energy consumption before and after implementing artificial intelligence

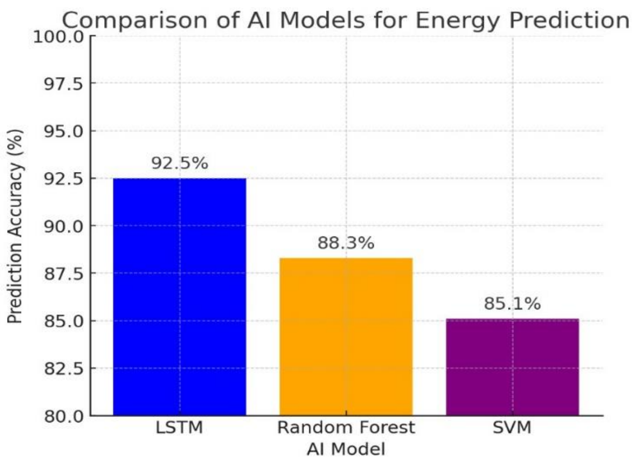


**Figure 2.** Bar chart comparing the accuracy of different AI algorithms in predicting energy consumption.

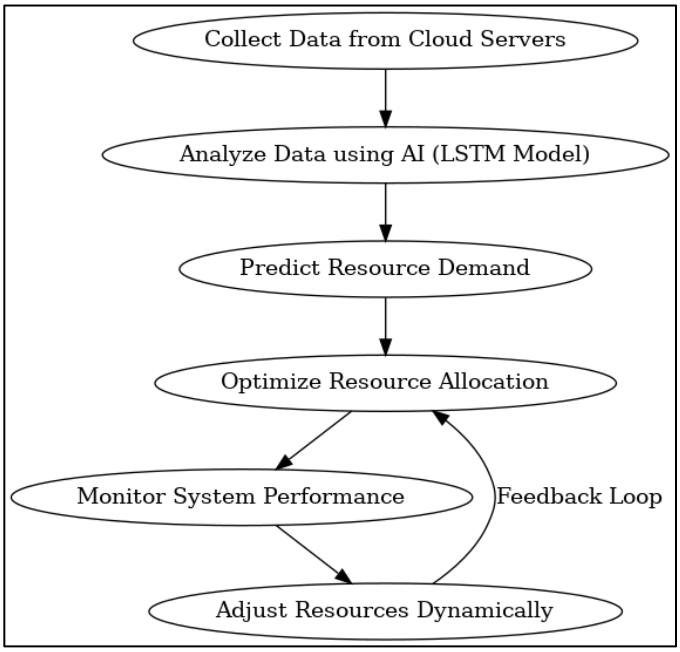


**Figure 3.** Flowchart illustrating the intelligent decision-making process in cloud resource management using AI.

## PREDICTIVE MODELING AND FORECASTING

Artificial Intelligence (AI) has experienced a renaissance and is implemented in various application scenarios. The feasibility of integration of AI technologies in cloud computing and management using the example of cloud infrastructure state management is explored. By contrasting different usage scenarios and matching AI technologies, opportunities are identified and the open problems are discussed. With the proliferation of cloud computing and data centers, management of infrastructure and applications running on cloud platforms is becoming increasingly complex and rate-limiting. Conventional mechanisms require users or system administrators to set adequate parameters or thresholds based on historical observations, but performance of the cloud infrastructure can unexpectedly deviate from expectations, for example in cases of sudden outages, performance regressions and resource contention. This often results in serious application performance degradation, poor resource utilization or even the failure of the entire cloud platform. On the other hand, the vast magnitude of monitoring data progressively prohibits human operators from observing the full state of cloud infrastructure in real time and immediate responsiveness is imperative to prevent severe performance degradation. Consequently, automated mechanisms and new technologies are needed to manage the state of cloud infrastructures and enable timely and adequate actions [23].

## CHALLENGES AND FUTURE DIRECTIONS

Cloud Computing is experiencing rapid growth and various researchers are actively working to address the research challenges that arise within the Cloud Computing domain. Cloud Computing has evolved as a tool for solving complex challenges faced by researchers in Earth Sciences, in terms of computations and analyses. In the ongoing decades, advances in everything have been associated with Cloud Computing infrastructure [1]. Earth Scientists have faced limitations in handling big volumes of data due to inadequate available computation infrastructure; they've been forced to downscale data or curtail the analytical process in some manner to expedite simulations and processing. Cloud Computing presents another way of operating where the physical hardware is invisible with scalable resources and costs based on usage. With the augmentation of fog computing technology, micro-servers have been introduced at the edge of the network. Fog Computing (FC) enables devices at the network edge to outsource commands run on a centralised server.

## TECHNICAL CHALLENGES AND IMPLEMENTATION HURDLES

Despite its many potential benefits and widespread adoption, the implementation of cloud computing raises many technical challenges. Managing a large-scale cloud computing environment is a complex task [24]. Many high-level cloud management systems have been proposed. However, they are not practical and don't perform well in large-scale environments. In addition, most existing cloud management systems are built for a specific cloud infrastructure platform and are difficult to extend to other cloud platforms. The proposed design is a general, extensible, and scalable cloud management architecture for a large-scale cloud computing environment [25].

## ETHICAL CONSIDERATIONS AND DATA PRIVACY

Cloud Computing and Management (CCM) services based on the integration of Artificial Intelligence (AI) technologies are constantly increasing. In particular, CCM systems have been proposed that are able to exploit AI technologies, such as data mining or simulation technologies, to provide a feedback of optimization of their functioning and of provisioning of their services. Furthermore, such systems might be easier to use for non-specialized users and might reduce the costs associated to the hiring and/or education of proper personnel. The use of the CCM might potentially raise concerns about data privacy since a leakage of the CCM configuration status and/or operating logs would disclose private/critical information. It is natural to a designer to investigate solutions to ensure the security and privacy of the data collected and disseminated by the CCM and to protect the CCM itself from attacks. A novel privacy approach based on the public cloud utilization, with sensitive data never exiting the client domain, is proposed. Such an approach is particularly useful when the use of the public cloud for backups or data processing is considered. Since the public cloud provider is not required to trust the client, an integrated cloud proxy for data sanitization and transmission is also provided. An analysis of the current status of the security guarantees to the public cloud clients is also performed, indicating the need for more efficient techniques [26].

## THE FUTURE OF AI IN CLOUD COMPUTING

The first research is going to carry out reliability analysis of the Algorithm Shepherding Systems in cloud computing based on the adaptation of the most significant models of the Software Architecture Analysis Method toolkit. The research is going to investigate the reliability models applied to basic ASS components through a process of data gathering and analysis, and the calculated results are going to be used to investigate how the models can be linked to provide an end-to-end solution [27] [28]. Empirical data is driving the software architecture analysis toolkit towards increasingly concrete models of reliability, safety, performance and security. These models are particularly suitable for systems with a high level of design detail, which have previously been constructed to the point of prototyping or implementation. A family of cloud computing based systems, whose architectural details are significant and experimentally determinable to build a simulator, is being worked on. However, system design has been conducted within the context of existing literature, common practices, and the measurement of some real world effects of an experimental service. For these reasons, it is not possible to fruitfully apply the monotonic, concrete safety model described in the toolkit [29].

## EMERGING TRENDS AND INNOVATIONS

Cloud data centers (CDCs) involve a wide range of computer resources including those related to the storing and processing of superfluous big data, and those for pampering to the strict query times required by online services. In view of the replication of data and redundancy from data centers (DCs), the building and operation of the CDCs becomes highly expensive [1]. With the ever-increasing proliferation of CDCs, the necessity for monitoring, governing, and scaling the performance of the CDCs raises compelling call for command language in predict and advance capacity planning scenarios. This issue poses a number of unique un-resolved complications due to dissimilar financial models and infrastructure. With the sector's trade air force globalization. CDCs data centers are necessary for providing storage resources and partially enabling services for many Internet and cloud providers, but the infrastructure complicatedness, size, and cost of a CDC can prohibit entry for young markets and new businesses seeking to display their large everyday applications [30]. To help ameliorate the financial burden for these eager providers, large cloud companies have introduced offering infrastructure as a cloud services. Given this new feature, one company can act not only as a CDC owner and operator but or else sublet the hardware infrastructure upon which a client organization can construct and manage their personal information technology services.

## RECOMMENDATIONS FOR FUTURE RESEARCH AND PRACTICE

Today, cloud computing providers require massive data centers containing thousands of servers to provide sufficient computing resources to an extensive user base. A Cloud Data Center (CDC) for delivering high numeracy services as well as to optimize their own operational costs combines distributed and virtualized servers and network gear. The granularity of data for monitoring the components in these numbers and size increases dramatically, and therefore, the number of big data monitoring concise statistics, arriving at an average of hundreds of statistics per host per minute compared with just a few performance metrics traditionally looked at, which cover several minutes or hours. The goal of this massive trend analysis is to detect performance regressions early, i.e., when they cure, being able to precisely identify the root cause (most probably one of the many thousands of parameters in the data centers) and be able to provide a repair action that is often non-trivial and not obvious from the analysis. However, as technology evolves, and applications change over time, what is normal operation for a complex distributed CDC is a moving target which makes it hard for fixed monitoring systems to remain effectiv [31] [32]. Yotascale is an automatic cloud performance monitoring tool able to learn behavioral models from historical data and detect problems via a suite of time series and machine learning techniques that it comes with out-of-box providing prediction features by utilizing historical series supporting structured and ad-hoc data exploration including automated ensemble models that capture the timing and correlation between multiple events at distinct levels of granularity and scale. Automatic alerts enable early problem detection that cannot always be diagnosed by viewing the raw data logs. A second set of tools enables the operator to visually correlate events and see the context around detection using on-demand log back-filling, browse back-in-time of dataset(s), and add custom annotations frequently by dividing the total amount of data into a timestamp date domain, network domain, and pod domain. Subsequent deep learning models such as autoregression, RNN, and LSTM take up the model development responsibility to identify the glaring abnormalities helping to save an event/crisis which in the absence of Yotascale could be catastrophic in nature. Finally, Yotascale grows models and insights end-to-end working with fellow SREs fabricating a higher-level of abstraction regarding what behaviors need to be preserved and accounted for across different causes and situations [1] [33]. Long term forecasts involve model complexity and training process continuously checking under and overfitting to support stable operation.

## CONCLUSION

SN is the abbreviation of scholarly name. Efforts have been made to keep the SN as it is translated, some terms may be modified. Before discussing the subject, it is important to discuss the terms of cloud computing and Artificial Intelligence (AI). SNIt is estimated that by 2040, about 95% of computing work will be conducted and controlled through the cloud, while 85% of business applications will be dependent on cloud services. The transformation towards a cloud-based environment is likely to introduce two key factors for innovation. On the one hand, the homogenization of infrastructure services and the globalization of the cloud computing market stimulate automation and integration in the management of cloud services and systems [3] [34]. Major economies such as the United States have already invested billions in R&D, including fundamental industries of oil and gas, military and airport management systems, and 76% of which have designated multistep AI implementation projects with controversial vision. At present, the development of AI in most enterprises is at a very early stage, approaching a mechanical imitation form of automation. At the same time, however, this makes consumption of computing resources too high. With the further development of technology in this field, AI is integrated more and more deep into these cloud systems to promote them to a more advanced platform.

## REFERENCES


[1] S. Gill, S. Tuli, M. Xu, I. Singh et al., "Transformative Effects of IoT, Blockchain and Artificial Intelligence on Cloud Computing: Evolution, Vision, Trends and Open Challenges," 2019.

[2] C. Fortuna, D. Mušić, G. Cerar, A. Čampa et al., "On-Premise Artificial Intelligence as a Service for Small and Medium Size Setups," 2022.

[3] S. S. Gill, S. Tuli, M. Xu, I. Singh et al., "Transformative Effects of IoT, Blockchain and Artificial Intelligence on Cloud Computing: Evolution, Vision, Trends and Open Challenges," 2019.

[4] T. Khan, W. Tian, and R. Buyya, "Machine Learning (ML)-Centric Resource Management in Cloud Computing: A Review and Future Directions," 2021.

[5] S. Singh Gill, M. Xu, C. Ottaviani, P. Patros et al., "AI for Next Generation Computing: Emerging Trends and Future Directions," 2022.

[6] V. Bucur and L. C. Miclea, "Multi-Cloud Resource Management Techniques for Cyber-Physical Systems," 2021.

[7] V. N. Tsakalidou, P. Mitsou, and G. A. Papakostas, "Machine learning for cloud resources management - An overview," 2021.

[8] Z. Xu, Y. Gong, Y. Zhou, Q. Bao et al., "Enhancing Kubernetes Automated Scheduling with Deep Learning and Reinforcement Techniques for Large-Scale Cloud Computing Optimization," 2024.

[9 M. J. A. Samah A Nasir, Zinah alhussein, “Prediction of Heart Disease Using Machine Learning Techniques”, IJRITCC, vol. 12, no. 2, pp. 91–95, Apr. 2024.
[10] M. J. A. Alkhafaji, A. Fadhil Aljuboori and A. A. Ibrahim, "Clean medical data and predict heart disease," *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey, 2020, pp. 1-7, doi: 10.1109/HORA49412.2020.9152870.
[11] M. J. A. Alkhafaji, “Development of a method for managing a group of unmanned aerial vehicles using a population algorithm”, EEJET, vol. 6, no. 9 (132), pp. 108–116, Dec. 2024.
[12] M. J. A. Alkhafaji, “A Review on Deep Learning-based Methods for Template Matching”, Vol. 20 No. 3 (2024).
[13] M. Canonico and L. De Russis, "A Comparison and Critique of Natural Language Understanding Tools," 2018.
[14] A. Pakmehr, A. Aßmuth, C. P. Neumann, and G. Pirkl, "Security Challenges for Cloud or Fog Computing-Based AI Applications," 2023.
[15] E. Papagiannidis, I. Merete Enholm, C. Dremel, P. Mikalef et al., "Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes," 2023.
[16] N. Chauhan, N. Kaur, K. Singh Saini, S. Verma et al., "A Systematic Literature Review on Task Allocation and Performance Management Techniques in Cloud Data Center," 2024.
[17] H. Alipour, "Model-Driven Machine Learning for Predictive Cloud Auto-scaling," 2019.
[18] M. J. Abed Alkhafaji et al., “the development of a method for increasing the reliability of the assessment of the state of the object.,” Eastern-European J. Enterp. Technol., vol. 131, no. 3, 2024.
[19] maytham kareem n. Alhasooni and mohammed jasim a. Alkhafaji, “advances and applications of template matching in image analysis”, Zenodo, Dec. 2024. doi: 10.5281/zenodo.14324241.
[20] mohammed jasim a. Alkhafaji, mohamad mahdi kassirand amir lakizadeh, “qatm-kcnn: improving template matching performance based on integration of convolutional neural networks and kalman filtering”, Zenodo, Dec. 2024. doi: 10.5281/zenodo.14325009.
[21] D. Ardagna, M. Ciavotta, R. Lancellotti, and M. Guerriero, "A Hierarchical Receding Horizon Algorithm for QoS-driven control of Multi-IaaS Applications," 1970.
[22] H. Adamu, M. Bashir, A. M. Bukar, A. J. Cullen et al., "An approach to failure prediction in a cloud based environment," 2017.
[23] B. Kang and H. Choo, "A cluster-based decentralized job dispatching for the large-scale cloud.," 2016.
[24] M. Ahmadi, M. Vali, F. Moghaddam, A. Hakemi et al., "A Reliable User Authentication and Data Protection Model in Cloud Computing Environments," 2015.
[25] S. Chintoju, "An Application for Decentralized Access Control Mechanism on Cloud Data using Anonymous Authentication," 2016.
[26] H. Fan, F. K. Hussain, M. Younas, and O. K. Hussain, "An integrated personalization framework for SaaS-based cloud services," 2015.
[27] O. Sohaib, M. Naderpour, and W. Hussain, "SaaS e-commerce platforms web accessibility evaluation," 2018.
[28] S. Kardani-Moghaddam, R. Buyya, and K. Ramamohanarao, "Performance-Aware Management of Cloud Resources: A Taxonomy and Future Directions," 2018.
[29] Z. Michalewicz and M. Michalewicz, "Machine intelligence, adaptive business intelligence, and natural intelligence," 2008.
[30] P. Agrawal, "Predictive Analysis for Cloud Infrastructure Metrics," 2019.
[31] K. A. Ahmat and H. Gobjuka, "Managing Clouds in Cloud Platforms," 2010.
[32] J. Sen, "Security and Privacy Issues in Cloud Computing," 2013.
[33] A. Gholami and E. Laure, "Security and Privacy of Sensitive Data in Cloud Computing: A Survey of Recent Developments," 2016.
[34] M. J. A. Alkhafaji, M. M. Kassir, and A. Lakizadeh, “A Review on Deep Learning-based Methods for Template Matching : Advancing with QATM-Kalman-CNN Integration,” vol. 10, 2025.