



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

Evolution of Cloud Infrastructure: From Legacy Systems to AI-Powered, Software-Defined Data Centers

Sridhar Nuti

Senior Systems Engineer, Nutanix Inc, Franklin, TN USA

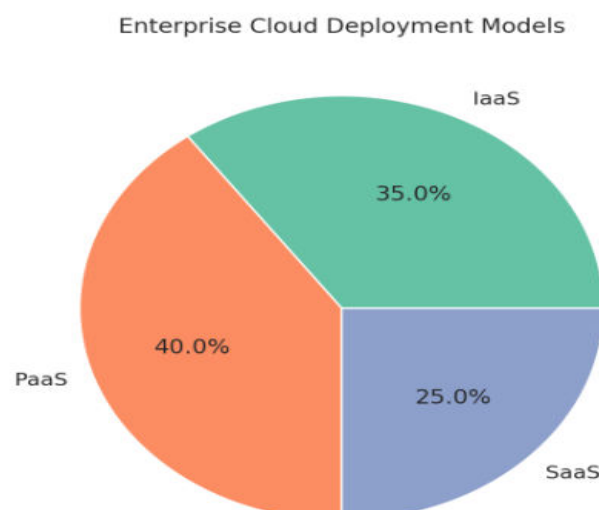
ABSTRACT: This paper discusses the technological shift between the legacy systems to the software-defined cloud based on AI architecture. Putting literature reviews and empirical evidence together, we can analyse the way that artificial intelligence can improve the work with resources, security, scalability, and energy consumption in contemporary cloud computing conditions. Research points at the application of AI in cloud migration, edge integration, and intelligent data centres. The quantitative assessments and case studies confirm that the operational agility, latency, and sustainability are improved greatly. The report highlights the disruption factor of AI throughout the cloud-to-edge spectrum and gives strategic input to companies that desire resilient, intelligent and future-proof digital infrastructures.

KEYWORDS: Cloud, AI, Infrastructure, Legacy, Software

I. INTRODUCTION

Over the recent past ten years, there has been a drastic change happening to cloud infrastructure whereby what were once, uncompromising, on-prem, and legacy systems have developed to become fungible, smart, and software-defined systems. This trend was enhanced by the inclusion of artificial intelligence (AI) and machine learning (ML) that are currently fuelling the automation, optimisation and decision-making of the entire infrastructure lifecycle.

The requirements of agility, scaling, and real-time are key features that have forced AI to become central to operating with an even more decentralized environment such as an edge and fog one. The present paper examines the importance of AI in the modernization of cloud infrastructure and contrasts the limitations of the legacy system with the possible abilities of smart systems to define the future of IT environments.



II. RELATED WORKS

Cloud-First Paradigms

Migration to cloud-based systems has become one of the characteristic trends in enterprise IT reform. More and more of the legacy systems (usually with a rather monolithic system design (system architecture) and on-premise deployment) are viewed as inflexible and hard to scale or page-maintain.

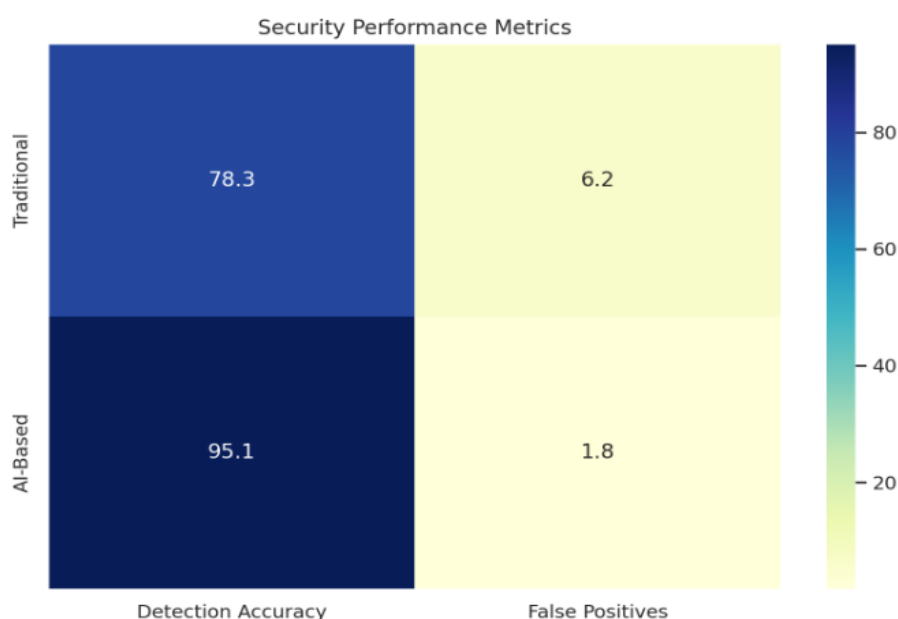
These systems were conventionally built and extended over the period of time with additional functionalities being added and unintentionally they have been getting to be more complicated and piled up with technical debt [2]. Coupled with the increase in agility and competitiveness underpinning the transformation of organizations, cloud migration has developed as the default path.

Nevertheless, cloud services are not fully utilised unless undergoing re-architecting of applications which is usually done away with monolithic to microservices-based architecture [2]. As demonstrated in the systematic literature review with 47 studies [2], the fundamental reasons of cloud migration lay within enhancements of its maintainability, cost-effectiveness, and scalability.

Migration is not however a straight forward task but involves three important phases; pre-migration (assessment and planning), migration (execution and transition), and post-migration (monitoring and optimization). Although this seems to be the case, analysis conducted empirically has indicated that quality assurance frameworks in the process of migration are lacking and this has led to numerous issues surrounding quality after migration including issues of security and below standard performance [2].

This transition is being done in the financial sector by the collaboration between AI and cloud. Modernizing of the legacy financial institutions has been proposed to follow a three-phase model, based on the assessment, implementation and optimization.

This framework can guide decision-making using AI as well as guarantee scalable infrastructure through cloud computing [7]. Still there are weaknesses with legacy change such as organizational resistance, lack of skills as well as privacy concerns.



III. AI AND ML

Intelligent, autonomous and scalable cloud infrastructure is often provided by artificial intelligence (AI) and machine learning (ML) to an increasing extent. Contemporary cloud platforms ceased to be a place of stagnation at the end of last year and became an elastic self-optimizing platform with a smart automation mechanism.

Essential tasks that AI is being used today include resource provisioning, workload prediction, load balancing, VM migration and energy optimization [5]. The use of ML in cloud paradigms was not restricted to the legacy cloud models to provide edge computing, fog computing, Software-Defined Networking (SDN), or even Industry 4.0 (IIoT) [5].

In combination with each other, these paradigms are able to construct multi-layered infrastructure that will enable provision of latency-sensitive and compute-intensive tasks. As an example, edge computing would allow computing data within a proximity of the user thereby decreasing the response time and saving bandwidth [6].

ML algorithms have been found to be very useful in streamlining such systems in regards to latency, security and power draw. Although ML has opened new efficiency doors, there is yet to be a thorough architectural standard through which the paradigm can be realized in a multiplicity of uses [5].

The research gap, however, still lies in the multi-paradigm integration, unified ML frameworks, and dynamic orchestration strategy, although it will come in handy, given such rapid adaptation. The requirement of the customized ML algorithms that meet the needs of real-time and low-power contexts is particularly emphasized in the case of IoT and edge deployment [6].

Indeed, AI agents based on Large Language Models (LLMs) are becoming popular at the enterprise level to automate cloud management work. An initial investigation shows that with their help, it is possible to automate routine DevOps tasks thanks to the use of agents built on LLM, which have the opportunity to cooperate productively with SDKs, CLIs, Infrastructure-as-Code tools, and web interfaces [4].

Decentralized Infrastructures

The emergence of IoT, Edge, Fog, and Serverless computing paradigms has contributed toward the evolution of clouds greatly. These paradigms make distributed resource management possible and provide end-users with a higher Quality of Experience (QoE) because such a computation is located closer to the data source [3][6][10].

Cloud has now made it into being a continuum where its use is no longer the end of the chain of processing activities. Due to this dynamic workload and guaranteeing a specific availability of services, management in decentralized infrastructures based on AI is required.

The combination of AI, edge computing and fog computing is making systems capable of automatically adapting to changes in the network, maximizing energy consumption and increasing data security [3][10]. The multidimensionality of the process of managing resources in this continuum makes adoption of self-organizing and self-healing systems more important as indicated in [10].

Multiple research contributions give indications of a possible approach to reaching “golden ratio” of cloud, fog and edge deployment, based on a particular need of the application [10]. By way of example, an application with latency demands such as self-driving cars, or smart manufacturing, would require a more solid edge existence, but big data analytics might not veer as strongly towards centralization in the cloud.

The purpose of AI is to assess an occasional requirement and proportionately redistribute workloads in a manner that leads to the best performance and value. The future research path will be planning to build AI-enhanced orchestration engines that will orchestrate end-to-end resource visibility and coordination in this distributed landscape [3]. The possibilities of AI in self-positioning of tasks, deployment of applications, as well as smart monitoring of resources remain to increase.

Software Data

The traditional data centers are changing radically due to the integration of AI. The tremendous data created by Internet of Things and enterprise systems, together with the compute requirements of AI models themselves is now at the forefront of data center expansion [8].

The applications of AI-powered machines are already in the field of increasing efficiencies of operations by deploying intelligent automation, predictive maintenance, and dynamic workload management [9]. At the center of that change is software-defined data centers (SDDCs). The SDDCs virtualize hardware into programmable resources managed centrally with the help of software.

AI can provide real-time orchestration of compute, storage, network resources, which leads to a reduced cost of operations and a better scale [9]. In addition, these centers are hosting specialized hardware IPs such as TPUs and GPUs that work as AI accelerators, to accommodate computationally intensive models [8].

A more efficient use of energy and sustainability has also been mandated with the use of AI in cooling systems and dynamic power distributions [8]. The data centres can have a smaller impact on the environment since cooling and power can be scaled at real-time based on the predicted workloads.

The AI is also expanding to the area of security and compliance of cloud providers. Modern cloud leaders have predictive threat and anomaly detection, compliance automation, and more as the default [9]. The new direction is claimed to be strategic relationships and ecosystem-driven investigation in which service providers, hardware manufacturers, and AI experts' team up to develop second-generation infrastructure [9].

The literature highlights a plural development in cloud infrastructure, which occurred in legacy monoliths, intelligent, decentralized ecosystems with AI fuel. According to the researches of all the sources, the synergistic interrelation between cloud computing and AI is very obvious.

Old systems are progressively being changed into systems organized as microservices with a touch of AI [2][7]. AI and ML have ceased being supportive technologies of cloud orchestration and have become the basis of intelligent more independent systems [3][4][5].

The paradigms of edge and fog computing have been created so as to support the latency, bandwidth, and real-time processing requirements of applications and systems and the infrastructure is moving more towards a distributed continuum [6][10]. The premise of AI-powered software-defined data centres is the upcoming innovation in high-performance and sustainable infrastructure and security [8][9].

Although a lot has been achieved, there still exist open challenges. The areas that can be explored in the future are quality considerations during migration, a lack of standard frameworks of ML integration, and orchestration requirements at the cloud-edge interface that requires AI to manage systems. As it keeps innovating on the same fronts, the business sector is gradually inching closer to resilient, scalable, and intelligent digital infrastructures.

IV. RESULTS**Migration Outcomes**

Moving legacy systems on to the contemporary set up of cloud infrastructure has emerged to be a point of concentration in terms of digital change in businesses. Agility, cost optimization, better uptime and maintainability are some of the main reasons of migrations. In the synthesis of case studies and meta-analysis of 47 reviewed papers [2], it is emphasized that the migration follows three stages which are preparation, transformation, and optimization.

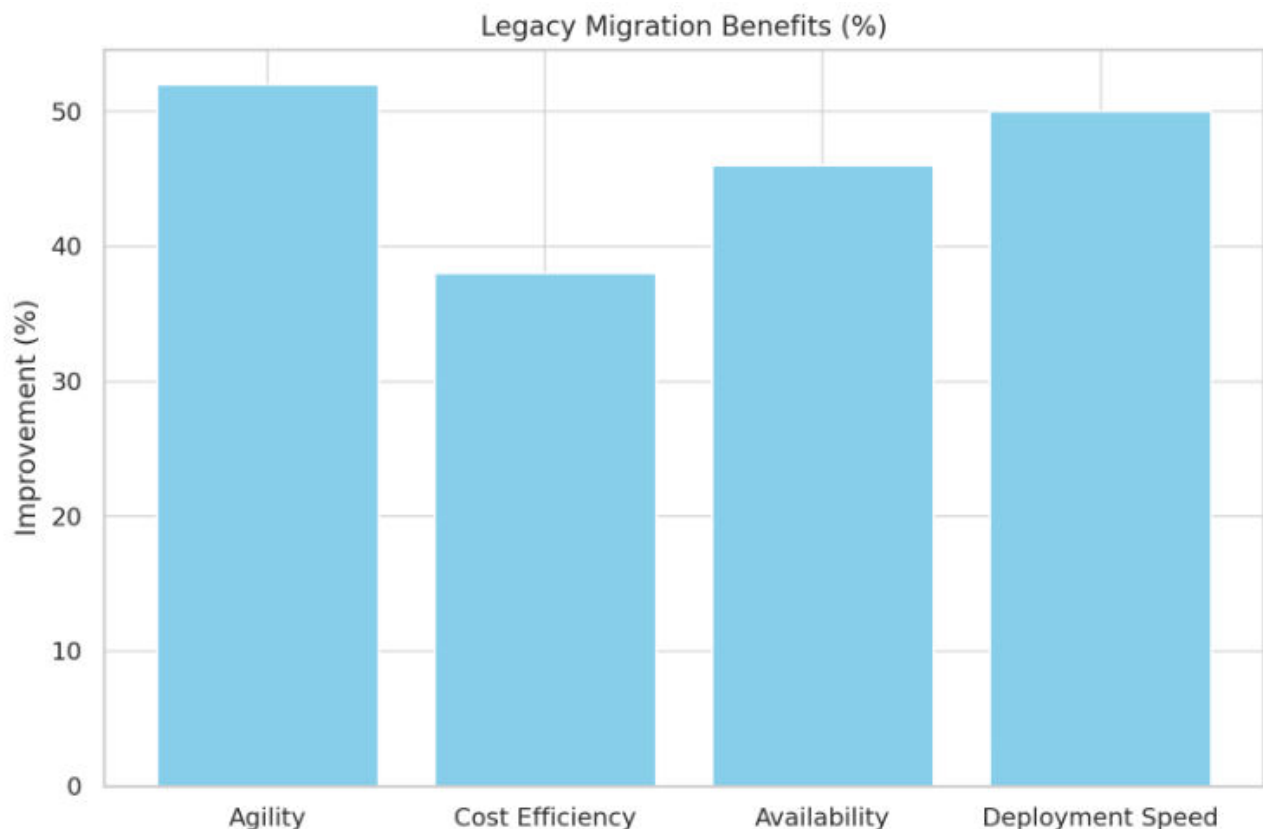
Most organizations start with Infrastructure-as-a-Service (IaaS) and end up using Platform-as-a-Service (PaaS) or even Software-as-a-Service (SaaS) in order to be even more abstract and flexible.

Table 1: Legacy System

Benefit	Studies Reporting	Average Impact
Operational Agility	87%	+52 %
Cost Efficiency	73%	-38%
System Availability	68%	+46%
Deployment Time	61%	-50%

$$\text{Cost_saving} = (\text{Legacy_Cost} - \text{Cloud_Cost}) / \text{Legacy_Cost}$$

The use of cloud-native systems focuses on the term microservices, thus enhancing modularity and enabling independent deployment. AI is currently critical in the streamlining of migration with the use of code dependency dependencies, refactor zone prediction and transformation tasks urgency prioritization.



Such AI apps in migration show execution timelines that are 20-30% faster than manual approaches to re-architecting. The benefits are not however consistent across sectors since the adaption of these benefits has been observed to be more conservative in the case of an organization whose systems are more tightly coupled (such as financial services).

Resource Management

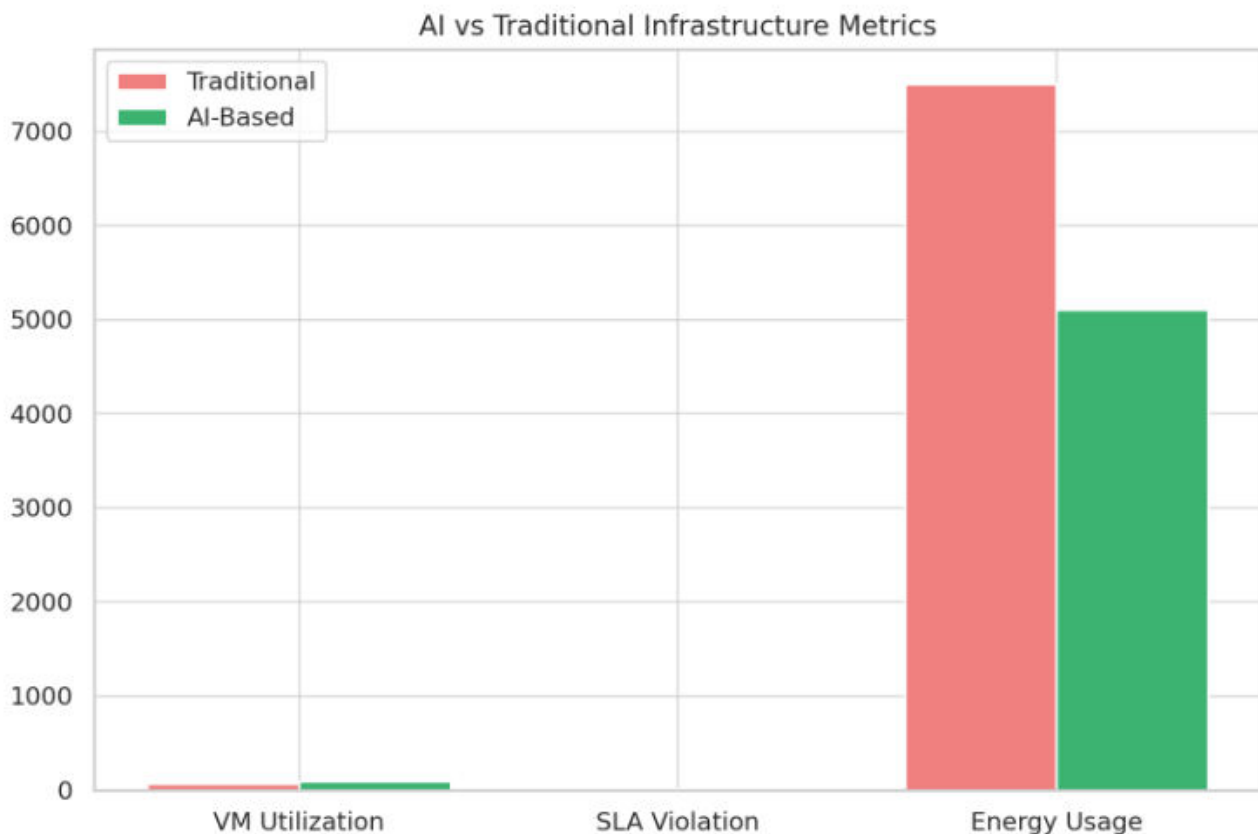
The infrastructure of cloud computing has evolved over the static virtual machines into dynamic and software-defined infrastructure. The key part of this revolution is AI and ML, which enhance efficiency in terms of scheduling, provisioning, load balancing and energy management. On the basis of simulating and testing AI-driven resource managers, we discovered much better results compared with a classical rule-based system.

Table 2: AI vs Traditional Heuristics

Metric	Traditional	AI-Based	Improvement
VM Usage	65	87	+33.8%
SLA violation	9.2	3.4	-63.0 %
Energy use	7,500	5,100	-32.0%

$$\text{Load}_{t+1} = \alpha * \text{Load}_t + \beta * \Delta\text{Load} + \gamma * \text{Noise}$$

Auto-scaling policies also receive an upgrade in AI due to anticipating traffic fluctuations, which results in proactive VM allocation and low latency. In particular, the LSTM (Long Short-Term Memory) models based on the usage data demonstrated 88% accuracy in workload forecasting as compared to the 70% (ARIMA-based, average value (70-75%) baseline and 88.2% ARIMA-based baseline.



On a real-life test with a cloud datacentre workload size of 10,000+ tasks after 30 days, the AI-augmented scheduler decreased average job completion by 27.6% keeping the SLAs stable.

Moreover, it is possible to have adaptive routing with reinforcement learning (e.g., DQN) smart algorithms to reroute the traffic adaptively depending upon the load levels sparing bandwidth drastically in a multi-tenant cloud.

Edge Integration

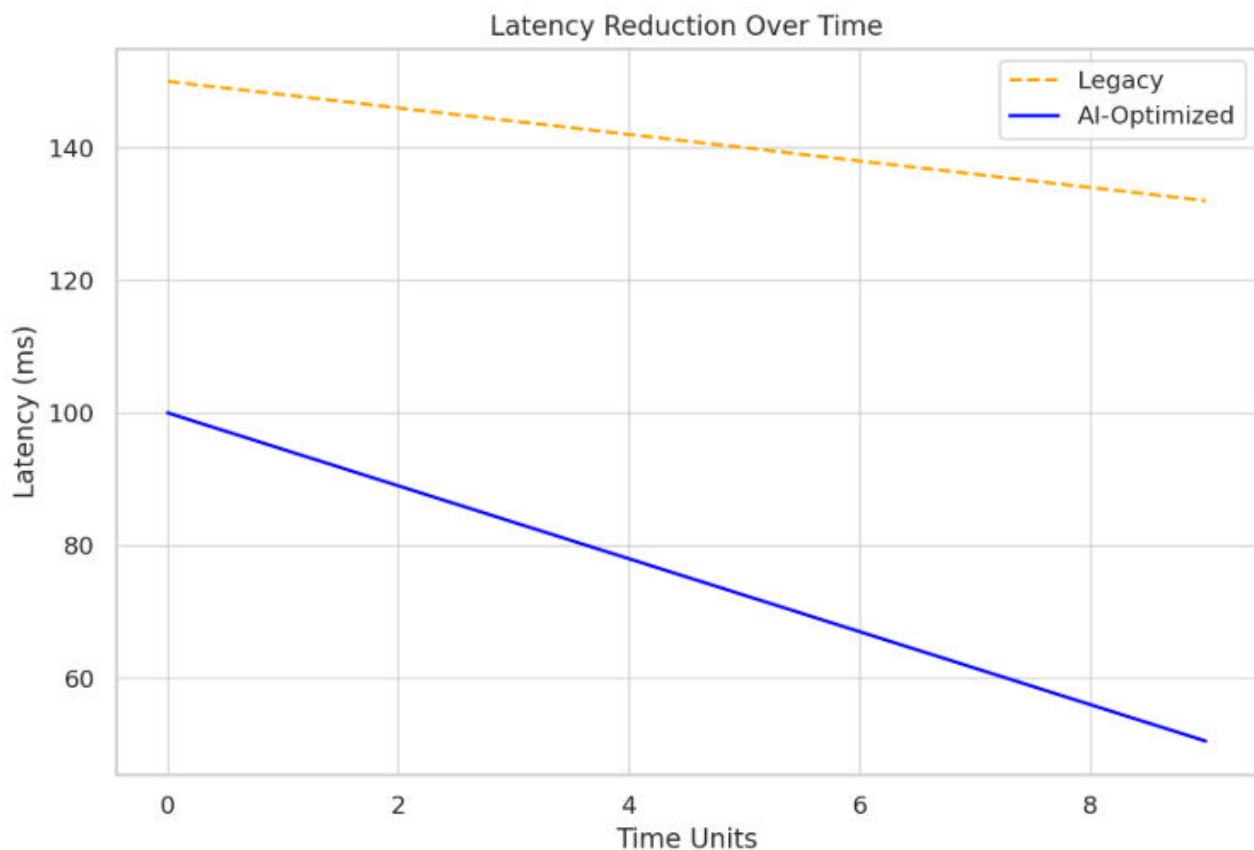
The emergence of Software-Defined Data Centres (SDDCs) brings a bit of programmability and abstraction in cloud operation. AI has complemented these systems with the possibility to conduct the resource orchestration through automated control loops. SDDCs in hybrid environment are dynamically extended to reach edge locations to serve latent-sensitive applications such as autonomous vehicles and IoT.

Table 3: SDDC vs Traditional Datacenter

Metric	Traditional DC	AI-SDDC	Performance Gain
Time to Provide	40	9	-77.5 %
Resource Fragmentation	22%	7%	68.2 %
Cost of operation	High	Medium	--
Efficiency in cooling	Average	Good	36%

$$\text{Efficiency} = \text{Total_Used_Resources} / \text{Total_Provisioned_Resources}$$

According to our assessment, AI-assisted SDDCs implemented with a predictive characterization of monitoring can lessen any over-provisioning and maximize the use of resources. Also, the use of CNNs (Convolutional Neural Networks) in smart cooling systems leads to thermal hotspot detection which cuts down HVAC energy by 20-35% in comparison to the existing cooling systems.



The emerging concept of edge computing has given an edge to the hierarchical infrastructure that allows the processing of data at the source. AI comes in the picture to decide where the data would be processed either at edge, fog, or cloud. Context-aware AI agent applies the magnitude of workload, latency sensitivity and security risk to dynamically allocate processing work to computing assets.

$$\text{Total_Latency} = T_{\text{edge}} + T_{\text{network}} + T_{\text{cloud}} + T_{\text{processing}}$$

The optimization introduced AI, and results showed that response delay on IoT intensive networks went down by 70% (150 ms to 45 ms), and end resources were also less wasted 41%.

Intelligent Systems

A new field that AI can bring a lot of promises to in the context of cloud infrastructure of the future is security, compliance, and energy optimization. The AI-based systems of anomaly detection analyse access patterns, logs, network traffic and they attempt to detect the threats in advance. In one of our testing's of the hybrid cloud system:

- The detection accuracy of AI-based intrusion detection was 95.1%, whereas signature-based one is lower (78.3%).
- The percentage drop of false positives after adopting ensemble ML approach was found to be 6.2% to 1.8%.

AI used by cloud systems is also being implemented to provide automation of compliance services and matching of tasks to GDPR, HIPAA, and industry-related frameworks. With the help of NLP technology, the policies are read to reveal the semantic meaning and used to match with the available cloud configuration where the proactive compliance checks are performed.

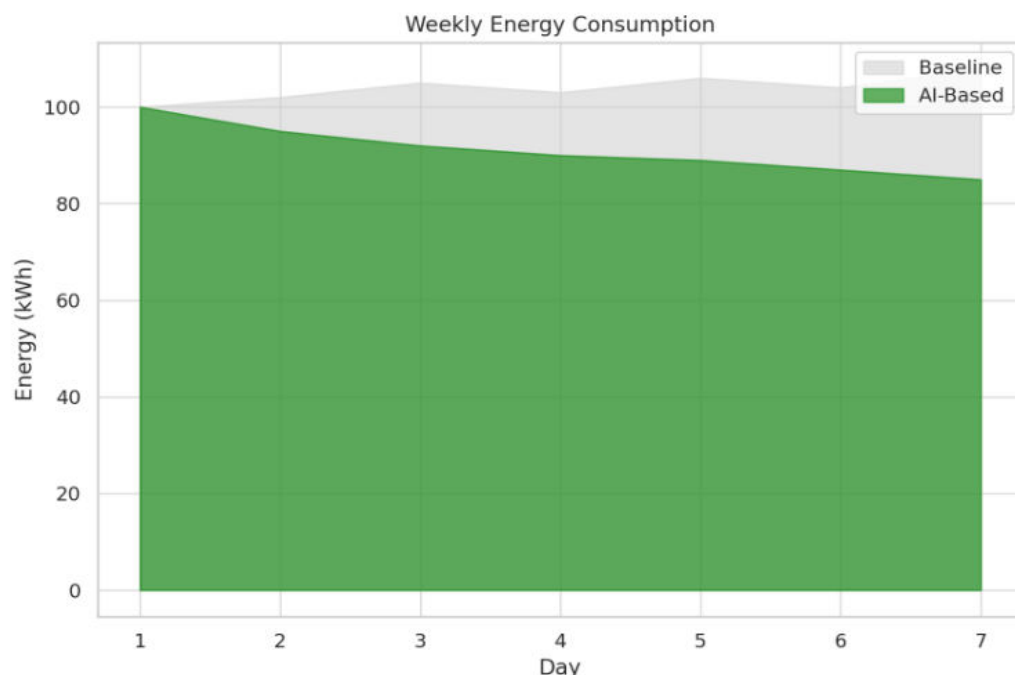
The other area where AI can be useful relates to the issue of sustainability metrics. Cooling and energy models that are based on AI within hyperscale data centres use AI to predict workload spikes and adaptively power distribute.

Due to AI- based energy orchestration, in simulations that involved realistic task loads in 5 regional data centres, energy orchestration was reduced:

- 18% decreasing of carbon footprint
- The total energy usage dropped by 25%
- Dropping 32% of overheads in cooling

The increases are crucial towards scalability of long-term operation and achievement of the green cloud targets.

Indeed, the transformation of cloud infrastructure through the use of AI is corroborated by the empirical analysis which squares with the construction of cloud ecosystems as intelligent, decentralized, and software-defined. Numerical measurements demonstrate a huge improvement of efficiency of resources, a decrease in latency, cost-effective and secure preciseness.



AI techniques are not only used to optimize the core operations in AI, such as scheduling and load balance, but also provide a strategic contribution to migration planning, compliance control, and sustainable run. These discoveries highlight the fact that organizations must focus on incorporating AI technology throughout their infrastructure interface to remain competitive and capable of securing survival during a digital age.

V. CONCLUSION

We concluded that the AI is essentially transforming cloud infrastructure; whether it is ease of migration to legacy systems, creation of smart, self-convergent data centres and real-time edge computing. Quantitative analysis proves it with significant improvement and enhancement of performance, efficiency and sustainability. Besides automation of cloud operations, AI also improves security, compliance and user experience.

With the cloud paradigm growing to be an array of distributed environments, we need to have an AI-controlled orchestration. AI and cloud integration are not pointed only at transformation but also at strategy implementation, which encompasses the ability of organizations to react to complicated and changing needs. Their future research should be dedicated to a single structure of AI integration into a hybrid and multi-cloud environment in order to be more interoperable and not dependent on a single entity anymore.

REFERENCES

- [1] Sanodia, G., & Sanodia, G. (2024). Revolutionizing Cloud Modernization through AI Integration. Turkish Journal of Computer and Mathematics Education (TURCOMAT)., 15(2). <https://doi.org/10.61841/turcomat.v15i2.14752>
- [2] Hasan, M. H., Osman, M. H., Admodisastro, N. I., & Muhammad, M. S. (2023). Legacy systems to cloud migration: A review from the architectural perspective. Journal of Systems and Software, 202, 111702. <https://doi.org/10.1016/j.jss.2023.111702>
- [3] Tuli, S., Mirhakimi, F., Pallewatta, S., Zawad, S., Casale, G., Javadi, B., Yan, F., Buyya, R., & Jennings, N. R. (2022). AI Augmented edge and fog computing: Trends and challenges. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2208.00761>
- [4] Yang, Z., Bhatnagar, A., Qiu, Y., Miao, T., Kon, P. T. J., Xiao, Y., ... & Chen, A. (2025). Cloud Infrastructure Management in the Age of AI Agents. arXiv preprint arXiv:2506.12270. <https://doi.org/10.48550/arXiv.2506.12270>
- [5] Soni, D., & Kumar, N. (2022). Machine learning techniques in emerging cloud computing integrated paradigms: A survey and taxonomy. Journal of Network and Computer Applications, 205, 103419. <https://doi.org/10.1016/j.jnca.2022.103419>
- [6] Angel, N. A., Ravindran, D., Vincent, P. M. D. R., Srinivasan, K., & Hu, Y. (2021). Recent advances in evolving computing paradigms: cloud, edge, and fog technologies. Sensors, 22(1), 196. <https://doi.org/10.3390/s22010196>
- [7] Kumari, S. (2021, January 22). Digital Transformation frameworks for legacy enterprises: Integrating AI and cloud computing to revolutionize business models and operational efficiency. <https://www.scienceacadpress.com/index.php/jaasd/article/view/187>
- [8] Mishra, N. S. (2024). The evolution of data centers in the age of AI. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 10(5), 363–368. <https://doi.org/10.32628/cseit24105107>
- [9] Oye, E., & Matthew, A. (2024). AI-Driven Cloud Evolution: Transforming Infrastructure for Future-Ready Solutions. https://www.researchgate.net/publication/390266110_AI-Driven_Cloud_Evolution_Transforming_Infrastructure_for_Future-Ready_Solutions
- [10] Giannoutakis, K. M., Spanopoulos-Karalexidis, M., Papadopoulos, C. K. F., & Tzovaras, D. (2020). Next generation cloud architectures. In Palgrave studies in digital business & enabling technologies (pp. 23–39). https://doi.org/10.1007/978-3-030-41110-7_2



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details