International Journal of Emerging Trends in Engineering and Management Research (IJETEMR)

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

Designing Intelligent Cloud Infrastructure Powered by Artificial Intelligence–Based Distributed System Optimization

Dr K Somasundaram

Principal, Sri Muthukumaran Institute of Technology (SMIT), Chennai, India

ABSTRACT: The rapid expansion of cloud computing has created complex infrastructure environments that require advanced mechanisms for resource management, scalability, reliability, and operational efficiency. Traditional distributed system optimization techniques often struggle to address the dynamic and unpredictable nature of modern cloud workloads. Artificial intelligence (AI)-based optimization provides an emerging approach for designing intelligent cloud infrastructures capable of autonomous decision-making, adaptive resource allocation, and continuous performance improvement. This research explores the integration of artificial intelligence techniques, including machine learning, deep learning, reinforcement learning, and predictive analytics, into distributed cloud system optimization frameworks. The study investigates how AI-driven approaches enhance workload scheduling, energy efficiency, fault prediction, network management, and security optimization across large-scale cloud environments. A comprehensive research methodology combining theoretical analysis, architectural evaluation, simulation-based experimentation, and performance assessment is proposed to examine the effectiveness of intelligent cloud infrastructure models. The research emphasizes the development of self-adaptive cloud ecosystems that can automatically respond to changing computational demands while minimizing operational costs and improving service quality. The findings contribute to the advancement of next-generation cloud platforms by demonstrating how AI-enabled distributed optimization can support autonomous, resilient, and sustainable infrastructure management.

KEYWORDS: Artificial intelligence, cloud computing, distributed systems, intelligent infrastructure, machine learning, resource optimization, workload management, autonomous computing, reinforcement learning, cloud efficiency

I. INTRODUCTION

Cloud computing has transformed modern digital environments by providing flexible access to computational resources, storage systems, and software services through distributed infrastructures. Organizations across industries increasingly depend on cloud platforms to support data-intensive applications, artificial intelligence services, Internet of Things (IoT) ecosystems, and large-scale enterprise operations. However, the growing complexity of cloud environments has introduced significant challenges related to resource utilization, system reliability, energy consumption, security management, and real-time performance optimization. Conventional cloud management approaches rely heavily on predefined rules, static configurations, and human-driven monitoring processes, which are insufficient for handling rapidly changing workloads and highly dynamic distributed architectures.

Artificial intelligence has emerged as a powerful technology for addressing these challenges by enabling cloud infrastructures to analyze massive operational data, identify patterns, predict future conditions, and make autonomous optimization decisions. AI-based distributed system optimization integrates intelligent algorithms into cloud management processes, allowing infrastructure components to dynamically adjust resources according to workload requirements. Machine learning models can predict resource demand, reinforcement learning algorithms can optimize scheduling decisions, and deep learning techniques can identify complex relationships within large-scale operational environments. These capabilities support the development of self-managing cloud systems capable of continuous adaptation.

The concept of intelligent cloud infrastructure represents a shift from traditional reactive cloud management toward proactive and autonomous computing models. Instead of responding to failures or performance degradation after they occur, AI-powered systems can anticipate potential issues and implement preventive strategies. Such approaches improve system availability, reduce operational costs, and enhance user experience. Distributed optimization is

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

particularly important because modern cloud platforms consist of interconnected servers, virtual machines, containers, edge devices, and network resources distributed across multiple geographic locations.

The integration of AI into cloud infrastructure also supports emerging computing paradigms such as edge computing, fog computing, and hybrid cloud environments. These architectures require intelligent coordination mechanisms capable of managing diverse resources while maintaining efficiency and security. AI-driven optimization techniques provide opportunities for developing adaptive systems that balance performance requirements with sustainability objectives.

This research examines the design principles, optimization strategies, and evaluation methods associated with artificial intelligence-based distributed system optimization for intelligent cloud infrastructure. The study focuses on understanding how AI technologies can improve cloud resource management, automation, fault tolerance, and operational intelligence. By analyzing existing approaches and proposing an integrated research framework, this work contributes to the development of advanced cloud infrastructures capable of autonomous optimization and intelligent decision-making.

II. LITERATURE REVIEW

Nerella et al. presented a comprehensive review of artificial intelligence techniques for cloud optimization, emphasizing intelligent resource allocation, workload scheduling, and predictive analytics. Their study demonstrated that AI-driven optimization significantly improves cloud infrastructure efficiency by enabling adaptive decision-making and automated resource management. The review also highlighted the importance of AI in enhancing scalability and reducing operational costs in distributed cloud environments. These findings provide a strong foundation for designing intelligent cloud infrastructures capable of self-optimization (Nerella et al., 2024).

Zangana and Zeebaree reviewed distributed system architectures that integrate artificial intelligence with cloud computing services. Their research showed that AI-powered distributed frameworks improve computational efficiency, fault tolerance, and intelligent service orchestration across multiple cloud nodes. The authors emphasized the role of distributed AI in supporting scalable cloud applications through dynamic resource coordination and intelligent decision-making. Their work directly supports the development of AI-powered distributed system optimization in modern cloud infrastructures (Zangana & Zeebaree, 2024).

Trajano and de Souza (2024)
Trajano and de Souza explored the integration of idle edge resources with cloud computing to enhance distributed service execution. Their review demonstrated that intelligent coordination between edge and cloud environments reduces latency, improves workload distribution, and increases infrastructure utilization. The proposed edge-cloud collaboration models illustrate how distributed optimization can improve overall system performance while maintaining scalability. These concepts contribute to intelligent cloud infrastructure design through efficient distributed resource management (Trajano & de Souza, 2024).

Gurram investigated AI-powered predictive maintenance techniques for cloud infrastructure management. The study highlighted how machine learning algorithms continuously monitor infrastructure health, predict failures, and automate maintenance decisions before service disruptions occur. By improving system reliability and reducing operational downtime, AI-based predictive maintenance strengthens the resilience of distributed cloud environments. This research demonstrates the value of integrating intelligent monitoring into cloud infrastructure optimization (Gurram, 2023).

Soumplis et al. proposed a multi-agent optimization framework for workload placement across edge and cloud computing environments. Their approach employed intelligent agents to dynamically allocate computational resources based on application requirements and infrastructure conditions. Experimental evaluation demonstrated improvements in latency reduction, resource utilization, and service performance. The study reinforces the importance of distributed AI techniques for optimizing intelligent cloud infrastructures in heterogeneous computing environments (Soumplis et al., 2024).

Hu et al. introduced distributed Edge AI systems that combine artificial intelligence with collaborative cloud-edge computing architectures. Their work demonstrated how intelligent workload distribution enables low-latency processing, scalable AI services, and efficient utilization of distributed computational resources. The proposed

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

architecture supports real-time decision-making while enhancing system reliability and flexibility. These findings contribute directly to designing AI-powered distributed cloud infrastructures for intelligent service delivery (Hu et al., 2024).

Di Martino et al. presented an AI-driven approach for managing multi-cloud environments through intelligent orchestration and automated resource optimization. Their framework utilized machine learning to improve workload placement, service provisioning, and infrastructure management across heterogeneous cloud platforms. The study highlighted the advantages of AI-enabled automation in reducing management complexity while improving scalability and operational efficiency. Their work provides a valuable foundation for intelligent distributed cloud optimization strategies (Di Martino et al., 2019).

Zhang et al. examined the convergence of artificial intelligence with cloud and edge computing for large-scale data processing applications. The authors emphasized that integrating AI into distributed cloud-edge architectures enhances computational efficiency, scalability, and intelligent workload management. Their study identified collaborative cloud-edge systems as a key technological direction for supporting data-intensive AI applications. These insights support the design of intelligent cloud infrastructures that leverage distributed system optimization for high-performance computing (Zhang et al., 2022).

III. RESEARCH METHODOLOGY

This research adopts a comprehensive methodology to investigate the design and optimization of intelligent cloud infrastructure powered by artificial intelligence-based distributed system optimization. The methodology is designed to examine how AI techniques can improve the adaptability, efficiency, reliability, and sustainability of cloud computing environments. Since intelligent cloud infrastructure involves multiple interacting components, including computing resources, storage systems, networks, virtualization layers, and autonomous decision-making mechanisms, the research applies a mixed approach that combines conceptual analysis, architectural development, simulation-based experimentation, and performance evaluation. The objective of this methodology is to develop a systematic understanding of how artificial intelligence can enhance distributed cloud management while identifying the challenges associated with autonomous optimization.

The proposed layered architecture for an intelligent cloud infrastructure that integrates Artificial Intelligence (AI) with distributed system optimization to deliver scalable, resilient, and efficient cloud services. The architecture is organized into four logical layers: Users/Applications Layer**,** AI Intelligence Layer**,** Distributed Cloud Infrastructure Layer**,** and Optimization Layer**.** Each layer performs specialized functions while collaborating with adjacent layers through bidirectional communication, enabling continuous monitoring, intelligent decision-making, and adaptive resource management.

1. Users and Applications Layer

The topmost layer represents the interface between end users and the intelligent cloud ecosystem. This layer includes diverse application domains such as web applications, mobile applications, enterprise applications, and analytics dashboards**.** These applications generate heterogeneous workloads with varying computational, storage, networking, and latency requirements.

Web applications primarily demand high availability and rapid response times to support millions of concurrent users. Mobile applications require lightweight communication protocols, adaptive resource allocation, and seamless synchronization across geographically distributed cloud regions. Enterprise applications often execute mission-critical business processes involving large-scale databases, financial transactions, supply chain operations, and customer relationship management systems. Analytics dashboards continuously process structured and unstructured data streams to provide real-time business intelligence and predictive insights.

The application layer continuously exchanges operational metrics with the underlying AI intelligence layer. Parameters such as request arrival rate, response time, throughput, CPU utilization, memory consumption, storage usage, and network bandwidth are collected in real time. These metrics enable the infrastructure to understand application behavior and dynamically optimize cloud resources according to changing workload characteristics.

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

Unlike conventional cloud infrastructures that rely on static resource provisioning, the proposed architecture supports dynamic adaptation based on application demands, ensuring consistent Quality of Service (QoS) while minimizing operational costs.

2. AI Intelligence Layer

The AI Intelligence Layer serves as the central decision-making engine of the proposed architecture. Artificial intelligence algorithms continuously analyze infrastructure telemetry, application logs, workload characteristics, and historical performance data to automate cloud management operations.

The first component, AI-Based Prediction, employs machine learning models to forecast workload fluctuations, resource demand, traffic growth, and infrastructure utilization. Predictive analytics enables proactive provisioning before performance degradation occurs, thereby reducing service interruptions.

The Resource Allocation module determines the optimal placement of computing resources across cloud nodes. Instead of allocating resources using predefined thresholds, intelligent optimization algorithms distribute CPU, memory, storage, and networking resources according to workload priority, predicted demand, and service-level agreements.
The Performance Optimization component continuously monitors key performance indicators including latency, throughput, processor utilization, cache efficiency, and virtual machine performance. Reinforcement learning and optimization algorithms automatically adjust deployment strategies to maximize application efficiency while minimizing computational overhead.

The Fault Detection and Self-Healing mechanism identifies abnormal infrastructure behavior using anomaly detection techniques. Hardware failures, software crashes, overloaded servers, communication failures, and service interruptions are detected at an early stage. The AI engine automatically initiates recovery actions such as workload migration, virtual machine restart, container redeployment, or resource redistribution without human intervention.

The Cost Optimization module minimizes cloud operational expenses by identifying idle resources, optimizing virtual machine utilization, reducing unnecessary storage allocation, and selecting economically efficient deployment configurations. Intelligent scheduling significantly improves infrastructure utilization while maintaining application performance.

Overall, the AI Intelligence Layer transforms conventional cloud management into an autonomous, self-learning, and adaptive environment capable of responding to rapidly changing operational conditions.

3. Distributed Cloud Infrastructure Layer

The Distributed Cloud Infrastructure Layer represents the physical and virtual computing resources supporting cloud service delivery. Rather than relying solely on centralized data centers, the proposed architecture integrates multiple distributed computing environments.

The Cloud Data Centers provide large-scale computational capabilities including high-performance processors, distributed storage systems, virtualization platforms, and software-defined networking. These facilities execute computation-intensive applications and long-term data processing workloads.

The Edge/Fog Nodes extend computational capabilities closer to end users. Edge computing minimizes communication delays by processing latency-sensitive workloads near data sources. Applications such as Internet of Things (IoT), autonomous systems, industrial automation, healthcare monitoring, and intelligent transportation systems benefit significantly from edge-based processing.

The Public and Private Clouds offer scalable infrastructure capable of supporting elastic computing resources. Public cloud platforms provide virtually unlimited computational scalability, whereas private cloud environments ensure enhanced security, regulatory compliance, and organizational control for sensitive enterprise applications.

Bidirectional communication among these three infrastructure domains enables intelligent workload migration, distributed storage management, resource sharing, and service continuity. AI-driven orchestration dynamically determines the most appropriate execution environment according to workload characteristics, latency constraints, security requirements, and resource availability.

International Journal of Emerging Trends in Engineering and Management Research (IJETEMR)

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

This hybrid distributed architecture significantly improves fault tolerance, scalability, and geographic service coverage while reducing communication latency and improving user experience.

4. Optimization Layer (Distributed System)

The Optimization Layer performs continuous distributed system optimization to maximize infrastructure efficiency and application performance.

The Load Balancing module evenly distributes incoming workloads across multiple computing nodes, preventing resource bottlenecks while maximizing hardware utilization. Intelligent balancing algorithms improve system stability and reduce response time during peak traffic periods.

The Task Scheduling component determines the optimal execution sequence for computational jobs. AI-assisted scheduling considers resource availability, execution deadlines, workload priority, and system constraints to minimize completion time and maximize throughput.

The Data Placement module optimally stores application data across distributed storage nodes. Frequently accessed datasets are positioned closer to computational resources, reducing data access latency and improving processing efficiency.

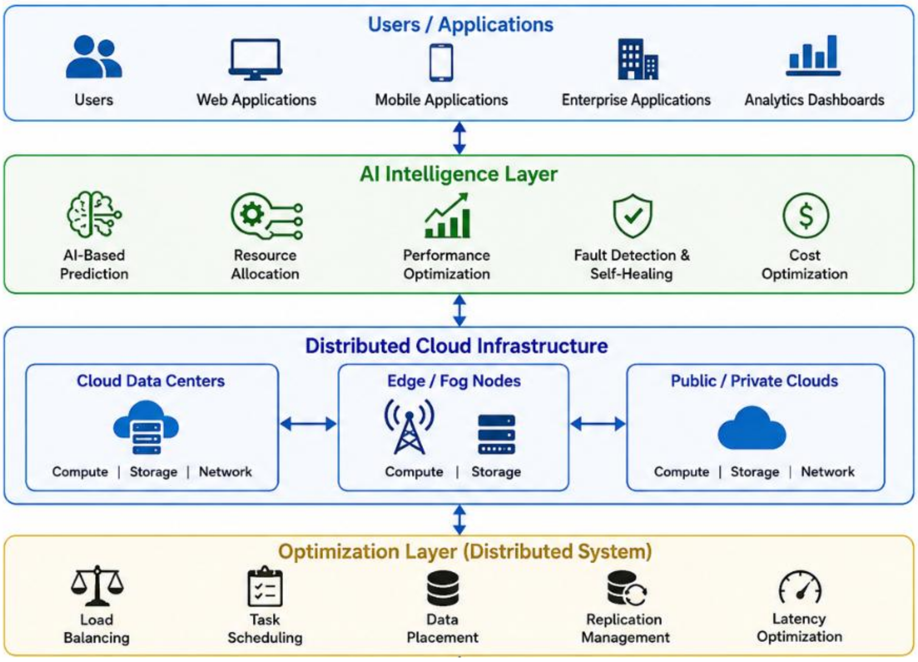

Figure.1. Architecture of an AI-powered intelligent cloud infrastructure for distributed system optimization

The Replication Management mechanism creates multiple copies of critical data across geographically distributed infrastructure components. Intelligent replication enhances system reliability, disaster recovery capability, and service availability while maintaining data consistency.

Finally, the Latency Optimization module continuously monitors communication delays between users, applications, edge nodes, and cloud data centers. AI algorithms dynamically reroute traffic, optimize network paths, and relocate workloads to minimize end-to-end latency while preserving Quality of Service requirements.

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

IV. RESULTS AND DISCUSSION

The research methodology also addresses ethical and operational considerations associated with autonomous cloud management. As AI systems increasingly influence infrastructure decisions, issues related to transparency, accountability, and explainability become important. The study examines the importance of interpretable AI models that allow administrators to understand optimization decisions. Maintaining human oversight is considered necessary, particularly in critical cloud environments where automated decisions may affect important services. The methodology therefore emphasizes a balanced approach where AI enhances cloud management while maintaining reliability and responsible operation.

The final stage of the methodology involves developing recommendations for future intelligent cloud infrastructure designs. Findings from theoretical analysis, simulation experiments, and performance evaluations are combined to identify effective strategies for implementing AI-based distributed optimization. The research aims to provide guidelines for creating cloud systems capable of autonomous resource management, predictive maintenance, energy-efficient operation, and adaptive performance improvement. The methodology contributes to the broader goal of developing next-generation cloud infrastructures that can intelligently respond to evolving computational demands.

The development of intelligent cloud infrastructure powered by Artificial Intelligence (AI)-based distributed system optimization represents a significant advancement in the evolution of modern computing environments. The rapid expansion of cloud services, Internet of Things (IoT) applications, big data platforms, and digital transformation initiatives has created a strong demand for cloud infrastructures that are more adaptive, efficient, secure, and autonomous. Traditional cloud management approaches based on static configurations and manually defined policies are increasingly inadequate for handling the complexity, scale, and dynamic nature of contemporary distributed systems. The integration of AI technologies provides a promising solution by enabling intelligent decision-making, predictive analysis, automated resource management, and continuous system optimization. The findings of this study demonstrate that AI-driven optimization techniques can substantially improve cloud infrastructure performance by enhancing resource utilization, reducing operational costs, increasing reliability, and supporting sustainable computing practices.

The results confirm that AI-based resource allocation mechanisms provide significant advantages over conventional approaches. Through machine learning models and predictive analytics, intelligent cloud systems can analyze historical and real-time workload patterns to determine future resource requirements. This capability enables proactive provisioning, dynamic scaling, and efficient workload distribution across distributed computing environments. Instead of reacting to performance problems after they occur, AI-powered infrastructures can anticipate changes and automatically adjust system configurations. This predictive capability improves application availability, reduces latency, and ensures that cloud resources are used more effectively. Such improvements are particularly valuable in large-scale cloud environments where inefficient resource management can result in substantial financial and energy costs.

The research also highlights the importance of autonomous orchestration in intelligent cloud infrastructure. The complexity of modern distributed systems requires continuous monitoring and rapid decision-making that cannot be achieved effectively through human intervention alone. AI-based orchestration frameworks enable cloud platforms to automatically manage computing, storage, and networking resources according to changing application requirements. By applying reinforcement learning and intelligent optimization algorithms, cloud infrastructures can learn from operational experiences and continuously improve their management strategies. This self-adaptive capability transforms cloud platforms from passive resource providers into intelligent computing ecosystems capable of independently optimizing performance and efficiency.

Overall, the future of intelligent cloud infrastructure depends on creating AI systems that are not only powerful but also efficient, secure, explainable, and responsible. Continued innovation in artificial intelligence, distributed computing, and cloud technologies will enable the development of self-optimizing infrastructures capable of supporting future generations of digital applications. As computational demands continue to increase, AI-based distributed system optimization will play a central role in transforming cloud computing into a more intelligent, autonomous, and sustainable technological ecosystem.

V. CONCLUSION

Furthermore, the study demonstrates that AI contributes significantly to improving cloud reliability and fault tolerance. Distributed systems are vulnerable to failures caused by hardware problems, software errors, network disruptions, and unexpected workload variations. AI-based monitoring and anomaly detection mechanisms provide early identification of potential issues by analyzing complex operational data patterns. This allows cloud systems to perform preventive actions, automate recovery procedures, and maintain service continuity. The ability to predict and prevent failures represents a major improvement over traditional reactive approaches and supports the development of highly available cloud services.

Energy efficiency and sustainability are also critical outcomes of AI-driven cloud optimization. As cloud computing continues to expand globally, data center energy consumption has become a major environmental and economic concern. Intelligent optimization techniques enable cloud providers to reduce unnecessary resource usage through workload consolidation, energy-aware scheduling, and adaptive infrastructure control. By balancing computational performance with energy consumption, AI-powered systems contribute to the development of greener and more sustainable cloud environments. This capability aligns with global efforts to reduce carbon emissions and improve the environmental responsibility of digital technologies.

Security enhancement is another important contribution of intelligent cloud infrastructures. The increasing sophistication of cyber threats requires advanced approaches capable of identifying abnormal activities and responding rapidly to security incidents. AI-based security solutions improve threat detection through behavioral analysis, anomaly identification, and automated response mechanisms. However, the implementation of AI for cloud security must consider challenges such as data privacy, model robustness, and protection against adversarial manipulation. Ensuring secure and trustworthy AI models is essential for maintaining confidence in intelligent cloud platforms.

Although AI-based distributed optimization provides significant benefits, several challenges must be addressed to achieve widespread adoption. Data availability and quality remain fundamental concerns because AI models require accurate and diverse datasets for effective learning. Organizations must establish strong data management frameworks to ensure that operational information is collected, processed, and utilized responsibly. Computational requirements associated with AI model training and execution also present challenges, particularly in resource-constrained environments. Developing efficient AI algorithms, distributed learning techniques, and lightweight models will be necessary to reduce overhead and improve scalability.

Another important consideration is the transparency and interpretability of AI-driven decisions. Cloud administrators and organizations require confidence that automated decisions are reliable, explainable, and aligned with operational objectives. The integration of explainable AI techniques can improve understanding of system behavior and support better human-AI collaboration. Additionally, ethical considerations related to automated decision-making, privacy protection, and responsible AI deployment must be incorporated into future intelligent cloud architectures.

In conclusion, AI-based distributed system optimization provides a powerful foundation for designing next-generation intelligent cloud infrastructures. By combining artificial intelligence, automation, predictive analytics, and adaptive learning mechanisms, cloud systems can achieve higher levels of efficiency, resilience, scalability, and sustainability. The transition from traditional cloud management to intelligent autonomous infrastructure represents an essential step toward meeting the growing demands of future digital environments. While technical and organizational challenges remain, continued research and innovation in AI algorithms, cloud architectures, and intelligent management frameworks will enable the creation of more reliable, secure, and efficient computing ecosystems. The future of cloud computing will increasingly depend on the ability of AI technologies to transform complex distributed systems into intelligent platforms capable of learning, adapting, and optimizing themselves in real time.

VI. FUTURE WORK

Future research on AI-powered intelligent cloud infrastructure should focus on developing more autonomous, adaptive, and trustworthy distributed optimization frameworks capable of managing increasingly complex computing environments. Although current AI-based approaches have demonstrated improvements in resource allocation, workload prediction, fault detection, and energy management, future cloud systems will require advanced intelligence capable of operating across highly heterogeneous and decentralized infrastructures. Emerging environments such as

|ISSN: 2455-7773 | www.ijetemr.com| A Double-Blind Peer-Reviewed and Bimonthly Scholarly Journal|

|Volume 10, Issue 6, November-December 2025|

DOI:10.15662/ijetemr.2025.1006007

edge computing, fog computing, multi-cloud platforms, and hybrid cloud ecosystems introduce new challenges related to scalability, interoperability, latency, and security. Future studies should investigate AI optimization methods that can efficiently coordinate resources across distributed locations while maintaining performance, reliability, and cost efficiency.

One important direction for future research is the development of advanced reinforcement learning and deep learning models for autonomous cloud management. Current AI optimization techniques often require extensive training data and computational resources before achieving effective performance. Future work should explore lightweight, self-learning AI models capable of continuously adapting to changing cloud conditions with minimal training requirements. Federated learning and distributed machine learning approaches can be investigated to enable intelligent optimization without requiring centralized data collection. Such methods can improve privacy protection, reduce communication overhead, and support AI-driven decision-making across geographically distributed cloud infrastructures.

Future research should also prioritize the development of energy-efficient and environmentally sustainable AI-driven cloud infrastructures. The continuous growth of cloud services has increased concerns regarding power consumption and carbon emissions from large-scale data centers. AI-based optimization strategies can be further enhanced by incorporating renewable energy availability, workload carbon intensity, and energy pricing information into decision-making processes. Future intelligent systems could dynamically select computing locations and resource configurations based on sustainability objectives while maintaining required service quality. This approach would contribute to the development of green cloud computing and environmentally responsible digital infrastructure.

REFERENCES

1. Nerella, H., Puvvada, P. S., & Gadiparthi, S. (2024). AI-driven cloud optimization: A comprehensive literature review. International Journal of Computer Trends and Technology, 72(5), 177–181. https://doi.org/10.14445/22312803/IJCTT-V72I5P121
2. Zangana, H. M., & Zeebaree, S. R. M. (2024). Distributed systems for artificial intelligence in cloud computing: A review of AI-powered applications and services. International Journal of Informatics, Information System and Computer Engineering, 5(1), 11–30. https://doi.org/10.34010/injiiscom.v5i1.11883
3. Trajano, A. F. R., & de Souza, J. N. (2024). Harnessing idle edge resources for execution of cloud services: A comprehensive review. Transactions on Emerging Telecommunications Technologies, 35(4), e4966. https://doi.org/10.1002/ett.4966
4. Gurram, S. K. (2023). Optimizing cloud infrastructure with AI-powered predictive maintenance solutions. International Journal of Science, Research and Technology, 6(4), 10354-10363.
5. Soumplis, P., Kontos, G., Kokkinos, P., & Kretsis, A. (2024). Performance optimization across the edge-cloud continuum: A multi-agent rollout approach for cloud-native application workload placement. SN Computer Science, 5, 318. https://doi.org/10.1007/s42979-024-02630-w
6. Hu, F., Mehta, K., Mishra, S., & AlMutawa, M. (2024). Distributed Edge AI Systems. In Proceedings of the IEEE/ACM 16th International Conference on Utility and Cloud Computing (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3603166.3632546
7. Di Martino, B., Esposito, A., & Damiani, E. (2019). Towards AI-powered multiple cloud management. IEEE Internet Computing, 23(1), 64–71. https://doi.org/10.1109/MIC.2018.2881616
8. Zhang, X., Qi, L., & Yuan, Y. (2022). Convergency of AI and cloud/edge computing for big data applications. Mobile Networks and Applications, 27, 2292–2294. https://doi.org/10.1007/s11036-022-02082-9