AI-Powered Cloud Cost Governance Systems: Automating Budget Management, Resource Allocation, and Policy Compliance for Optimized Cloud Expenditure

Charan Shankar Kummarapurugu

Senior DevOps Engineer — Cloud,
DevSecOps and AI/ML Brambleton, VA, USA
Email: charanshankar@outlook.com

Abstract
The rapid adoption of cloud computing by enter- prises has introduced significant complexities in managing and controlling cloud expenditure, especially in multi-cloud environ- ments. Traditional approaches to cloud cost governance often rely on manual interventions and static rules, leading to inefficiencies and unpredictable costs. To address these challenges, this paper proposes an AI-powered cloud cost governance system that auto- mates budget management, resource allocation, and policy com- pliance. The system leverages machine learning models to predict resource demands and costs, enabling real-time adjustments to cloud resources based on both historical data and current usage patterns. Moreover, AI-driven policy enforcement ensures that pre-defined budgetary constraints are adhered to automatically, reducing human intervention and improving cost efficiency. A simulation study demonstrates the effectiveness of the proposed system, showing significant reductions in cloud spending and improved utilization of cloud resources. The paper concludes by discussing the future potential of integrating reinforcement learning to further enhance cloud governance systems.

Keywords: AI in Cloud Computing, Cloud Cost Gover- nance, Policy Automation, Resource Allocation, Budget Manage- ment, Multi-Cloud Optimization

I. INTRODUCTION

Cloud computing has become an essential component of modern IT infrastructure, offering enterprises unprecedented flexibility, scalability, and accessibility. By leveraging cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud, organizations can scale their ap- plications and infrastructure on-demand without the need for large upfront investments in physical hardware. How- ever, the very features that make cloud computing attrac- tive—scalability and flexibility—also introduce significant challenges in managing costs. Without effective governance mechanisms, organizations often face unpredictable billing spikes, inefficiencies in resource utilization, and difficulties in enforcing budgetary policies [1].

Traditional approaches to cloud cost management rely heav- ily on manual processes or simple rule-based systems to monitor and control expenditure. While these methods can be effective in small-scale or static environments, they fall short in handling the dynamic and complex nature of multi-cloud and hybrid cloud infrastructures. Manual interventions and static rules are not equipped to respond to real-time changes in resource usage, pricing fluctuations, or sudden demand spikes, leading to either over-provisioning (wasted resources) or under-provisioning (performance degradation) [2].

In response to these challenges, AI-powered systems have emerged as a promising solution for cloud cost governance. By integrating machine learning (ML) models into the cloud management pipeline, organizations can automate the predic- tion of resource demands, optimize resource allocation, and enforce cost policies dynamically. These systems are capable of learning from historical usage patterns and making real-time adjustments based on current conditions, thereby improving cost efficiency and reducing manual intervention [3].

This paper presents a comprehensive framework for AI- powered cloud cost governance, focusing on three key aspects: automated budget management, intelligent resource allocation, and dynamic policy compliance. The proposed system lever- ages AI to continuously monitor cloud usage, predict future resource needs, and enforce cost policies. By automating these processes, the system ensures that cloud resources are used efficiently and that costs are controlled without compromis- ing on performance. The paper also presents a simulation- based evaluation of the system's effectiveness in a multi- cloud environment, demonstrating significant cost savings andimprovements in resource utilization.

The rest of the paper is structured as follows. Section II reviews related works on cloud cost management and AI-basedoptimization techniques. Section III outlines the proposed architecture and methodology for the AI-powered cost gov- ernance system. Section IV presents the results and analysis from the simulation study. Finally, Section V concludes the paper and discusses potential future directions, including the integration of reinforcement learning for even more adaptive governance solutions.

II. RELATED WORKS

Cloud cost optimization has been extensively researched due to the growing complexity of managing dynamic cloud environments and the escalating need to control expenditures in multi-cloud infrastructures. This section reviews various approaches from rule-based systems to advanced AI-driven solutions.

- **Rule-Based Systems:** Early cloud resource management relied on static rules and predefined thresholds, which, while effective in small-scale environments, often failed to adapt to the unpredictable nature of modern cloudworkloads, resulting in issues like over-provisioning and under-provisioning [4], [5].
- **Machine Learning and AI-Driven Solutions:** To ad- dress the limitations of rule-based systems, researchers have employed predictive analytics and machine learning, particularly time series forecasting models like ARIMA and LSTM, to predict resource usage and adjust alloca- tions dynamically [6], [7].
- **Reinforcement Learning:** Recent studies have explored reinforcement learning for cloud resource optimization, with RL agents learning to balance performance and cost effectively, outperforming traditional methods [8], [9].
- **AI-Driven Policy Automation:** Despite the advances in AI for cloud management, there remains a gap in integrat- ing these technologies with cost governance frameworks. AI-driven policy automation can enforce budgetary poli- cies in real-time, improving scalability and effectiveness in dynamic environments [10], [11].
- **Cloud Provider Tools:** Major cloud providers have integrated AI tools to assist with cost management, such as AWS's Cost Explorer and Azure's Cost Management + Billing. However, these tools often require manual oversight and do not fully automate cost governance [12].

In conclusion, while AI-driven cloud cost optimization has progressed, existing solutions typically focus on specific as- pects of cloud management. This paper proposes an integrated AI-powered cloud cost governance system that optimizes resource allocation and automates budget management and policy compliance, dynamically controlling expenditure acrossmulti-cloud environments.

III.PROPOSED ARCHITECTURE AND METHODOLOGY

The proposed AI-powered cloud cost governance system is designed to provide a comprehensive framework for au- tomating cloud cost management through budget enforcement, dynamic resource allocation, and real-time policy compliance. The architecture leverages state-of-the-art machine learning algorithms to predict future resource usage and optimize cloud expenditures across multi-cloud environments. By continu- ously monitoring cloud usage and applying AI-driven insights, the system dynamically adjusts cloud resources to meet work- load demands while ensuring adherence to budget constraints.

A. System Architecture Overview

The architecture of the proposed AI-driven cost governance system is modular and composed of several interconnected components, each responsible for a specific function. These components include:

- **AI-Driven Cost Prediction Model** for forecasting cloud resource usage and cost trends.
- **Dynamic Policy Enforcement Engine** that ensures cloud usage adheres to predefined budgetary and policy con- straints.
- **Real-Time Resource Monitoring and Optimization Engine** to track cloud resources and optimize their usage.
- **Multi-Cloud and Hybrid Cloud Support Layer**, al- lowing resource management across different cloud plat- forms.
- **Security and Compliance Monitoring Module** to en- sure that cost optimization does not compromise security and regulatory compliance.

The architecture is designed to operate in a cloud-agnostic fashion, supporting major cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). Figure 1 provides an overview of the system architecture, showing how each component interacts with the cloud environment and the AI models to provide a fully automated cost governance system.

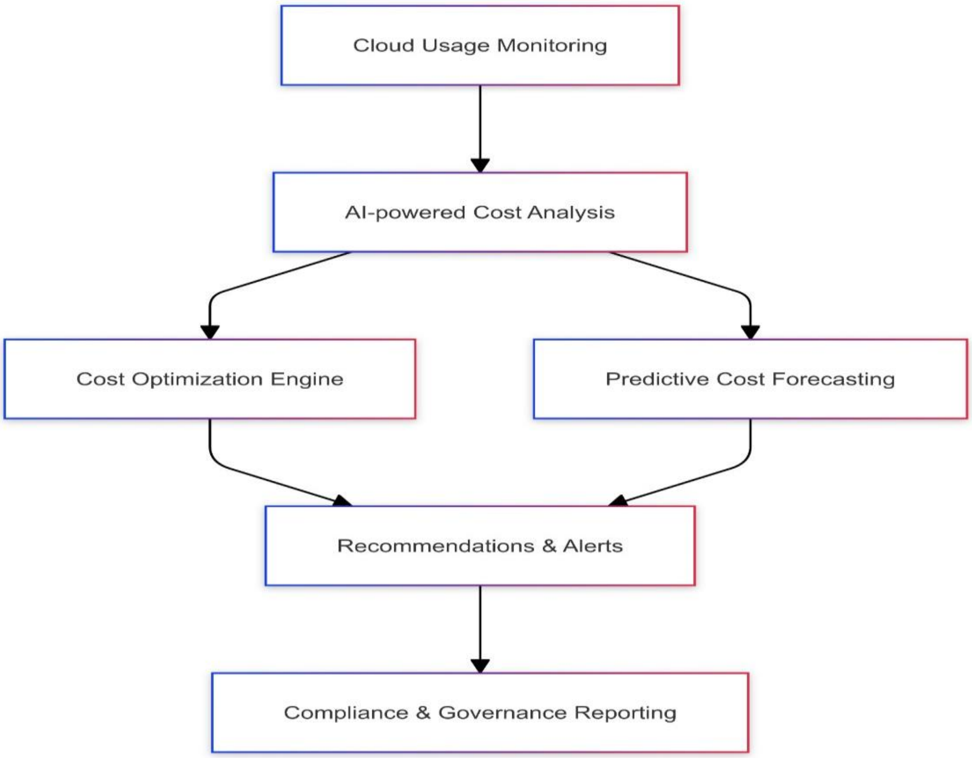

Fig. 1. Proposed AI-Powered Cloud Cost Governance System Architecture

B. *AI-Driven Cost Prediction Model*

The AI-driven cost prediction model is central to the architecture and plays a critical role in ensuring efficient resource utilization and cost control. The model employs machine learning algorithms to predict future resource usage based on historical data. The key objective of this module is to anticipate demand for cloud resources and estimate the associated costs, allowing for proactive adjustments to resource allocation before demand spikes occur.

1) Data Collection and Feature Engineering: The cost prediction model relies on data collected from multiple cloud services, including metrics such as:

- CPU utilization rates
- Memory consumption
- Network traffic
- Disk I/O
- Instance uptime and idle time

This data is gathered over time to create a rich dataset that captures usage patterns, workload fluctuations, and trends. Feature engineering is applied to preprocess this data, extract- ing meaningful features such as peak load periods, resource consumption during specific times of the day, and correlation between workload size and resource usage.

2) Machine Learning Models for Forecasting: The system uses advanced time series forecasting techniques to predict future resource usage. These models include:

- **ARIMA (AutoRegressive Integrated Moving Average)**: A classical statistical approach to time series forecasting that models the relationship between past and future values.
- **LSTM (Long Short-Term Memory Networks)**: A type of recurrent neural network (RNN) that is particularly ef- fective at capturing long-term dependencies in sequential data, making it ideal for predicting workload spikes and long-term trends.
- **Exponential Smoothing Methods**: For cases where workload patterns are relatively stable with minor varia- tions, simpler methods like exponential smoothing can be used to generate accurate forecasts with minimal computational overhead.

The AI-driven prediction model continuously updates itself with new data, ensuring that predictions remain accurate even as workloads evolve. By predicting resource needs in advance, the system allows for the adjustment of cloud instances and resources before bottlenecks or over-provisioning occurs. Fig- ure 2 illustrates the workflow of the AI-driven cost prediction process.

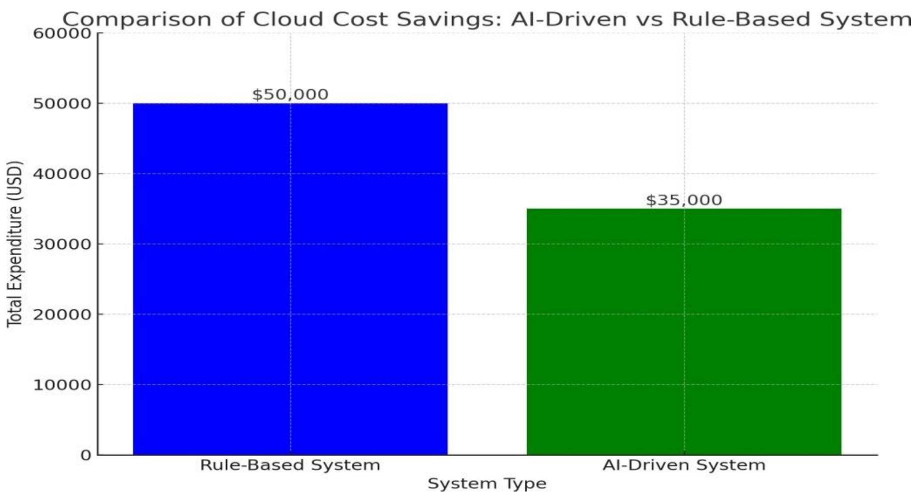

Fig. 2. AI-Driven Cost Prediction Model Workflow

C. *Dynamic Policy Enforcement Engine*

The dynamic policy enforcement engine ensures that cloud usage stays within predefined budgetary limits and follows organizational policies. It serves as the governance layer, using AI forecasts and real-time monitoring data to enforce costcontrols.

1) *Policy Definition and Enforcement:* Administrators de- fine policies based on budgets, resource usage thresholds, and service-level agreements (SLAs), such as:

- Monthly cloud expenditure limits.
- Resource usage caps (e.g., CPU utilization below 80%for non-critical workloads).
- Restrictions on deploying workloads in high-cost regions.

The engine monitors usage in real-time, comparing it against policies. Upon detecting violations, it triggers corrective ac- tions like scaling down resources, stopping non-critical pro- cesses, or migrating workloads to more cost-effective regions.

2) *Automated Decision Making and Remediation:* The en- gine uses AI to assess the cloud environment continuouslyand make decisions without manual intervention. Examples include:

- Automatically scaling down resources or shifting work- loads if regional usage exceeds thresholds.
- Decommissioning non-critical instances to prevent cost overruns.

By automating these processes, the engine minimizes hu- man error, enhances response times, and ensures continuous compliance with cost-related policies. Figure 3 illustrates the process.

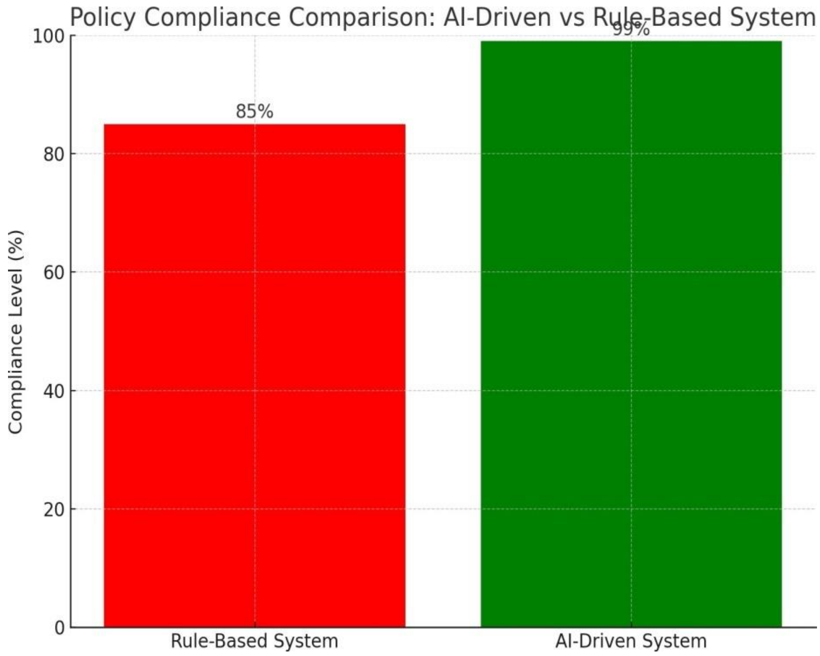

Fig. 3. Dynamic Policy Enforcement Process

D. *Real-Time Resource Monitoring and Optimization Engine*

The real-time resource monitoring and optimization engine is the operational core of the system, responsible for oversee- ing cloud resource usage and making timely adjustments basedon actual workloads. It leverages cost prediction models and policy enforcement engines to provide data-driven insights for effective cloud management.

- **Continuous Resource Monitoring:** Utilizes cloud provider APIs (e.g., AWS CloudWatch, Azure Monitor) to monitor resources in real-time, tracking metrics like CPU and memory usage, disk I/O, network bandwidth, and instance uptime. This enables the detection of anoma- lies, inefficiencies, or spikes in resource usage.
- **Optimization Algorithms:** Employs machine learning to analyze real-time data and optimize resource allocations through:

– **Consolidating underutilized resources:** Identifies and merges workloads from idle instances to enhance efficiency.
– **Resizing instances:** Adjusts instance sizes based on actual workload needs to minimize over-provisioning.
– **Dynamic workload shifting:** Relocates workloads across cloud regions or providers to capitalize on cost efficiencies.

This system's integration with AI and machine learning en- sures that cloud resources are managed efficiently, adapting dynamically to changes in workload demands and operational conditions. An AI-driven anomaly detection mechanism also identifies unusual resource usage patterns, triggering appropri-ate scaling actions or alerts to administrators as needed. The process is visualized in Figure 4.

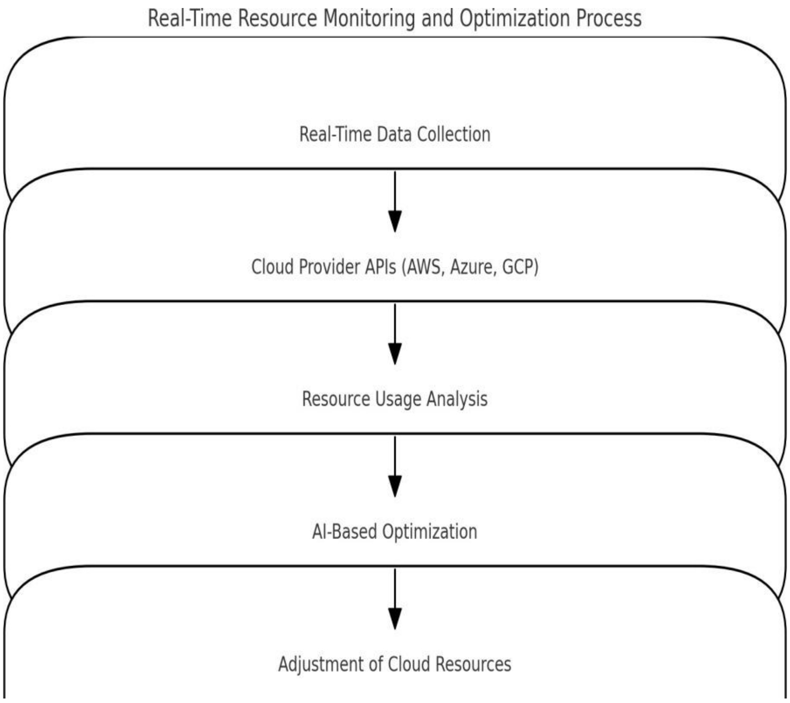

Fig. 4. Real-Time Resource Monitoring and Optimization Process

E. *Multi-Cloud and Hybrid Cloud Support*

The proposed system enhances resource management across various environments by supporting both multi-cloud and hybrid cloud configurations. Key features include:

- **API Integration:** Interfaces with APIs from major cloud providers such as AWS, Azure, and GCP to access usage data, predict costs, and enforce compliance policies.
- **Dynamic Workload Migration:** Facilitates the transfer of workloads to lower-cost or better-suited cloud regions based on real-time cost predictions, enhancing cost effec-tiveness.
- **Cloud-Agnostic Operation:** Allows optimization of re- source allocation across diverse cloud platforms, main- taining flexibility and compliance with internal gover- nance policies.

This approach ensures operational flexibility and cost effi- ciency while supporting diverse and complex cloud infrastruc-tures.

F. Scalability and Adaptability

The architecture of the proposed AI-powered system is built to scale seamlessly with an organization's expanding cloud infrastructure, capable of handling large-scale environments with thousands of instances. It uses AI models to analyze real-time data and make rapid adjustments to resource allocation. The scalability and adaptability features include:

- **Horizontal Scaling:** Adds monitoring agents and AI models across multiple regions or cloud providers, ac- commodating geographically dispersed deployments ef- fectively.
- **Elastic Workload Management:** Dynamically adjusts resource utilization based on workload demands, scaling up during peak times and scaling down during off-peak hours.
- **Real-Time Decision Making:** Leverages continuous data updates to make informed decisions on resource alloca- tion, policy enforcement, and cost prediction, enabling quick responses to changes in the cloud environment.
- **CI/CD Integration:** Integrates with CI/CD pipelines to embed AI-driven cost governance into the deployment process, ensuring that new cloud services are both cost- efficient and compliant with policies.

This system's cloud-agnostic capability ensures it operates effectively across various cloud platforms, such as AWS, Azure, and GCP, and in hybrid environments, making it suit- able for enterprises of all sizes and enhancing cost governance as organizations scale.

G. Security and Compliance Considerations

Security and compliance are critical in managing cloud resources, particularly in regulated industries such as finance, healthcare, and government. The proposed system ensures ro- bust security protocols and compliance monitoring to maintain integrity while governing costs.

- **Compliance Monitoring and Auditing:** The system includes a compliance monitoring module that adheres to regulations like GDPR, HIPAA, and PCI-DSS. Key functions include:
  - Ensuring data residency and sovereignty by control- ling where data is stored and processed.
  - Providing detailed audit logs and reports that en- hance transparency and accountability in resource allocations and policy enforcement.
- **Security Integration and Anomaly Detection:** Inte- grates with cloud-native security services such as AWS IAM, Azure Active Directory, and Google Cloud IAM to enforce access controls, encryption, and network security. AI-powered anomaly detection identifies and mitigates security threats promptly, with automated actions like access revocation and instance scaling.
- **Automated Incident Response:** The system's incident response capabilities can automatically correct compli- ance violations or misconfigurations in real-time. If sen- sitive data is stored incorrectly, it can rectify the error or alert administrators to intervene, thereby embedding se- curity checks directly into the cost governance workflow.

These features ensure that the system's optimization of cloud cost and resource usage does not compromise security or compliance, enabling organizations to dynamically enforce policies and handle cost control with predictive analytics and real-time monitoring across various cloud environments.

IV. RESULTS AND ANALYSIS

In this section, we present the results of evaluating the proposed AI-powered cloud cost governance system through a combination of simulations and real-world cloud usage data. The primary objective of the evaluation is to demonstrate the system's effectiveness in optimizing cloud expenditure, improving resource utilization, and automating policy en- forcement. We compare the performance of our system with traditional rule-based cloud cost management techniques and analyze the improvements in terms of cost savings, resource optimization, and policy compliance.

A. *Experimental Setup*

The evaluation was conducted in a simulated multi-cloud environment with resource usage patterns modeled after real- world enterprise cloud workloads. The simulated environment included:

- Cloud infrastructure from three major providers: AWS, Microsoft Azure, and Google Cloud.
- A variety of services, including virtual machines, storage services, and serverless computing instances.
- Workloads with varying demand patterns, including peri- odic spikes in resource usage and sustained high-demand periods.

The AI-powered system was configured with machine learn- ing models (ARIMA and LSTM) for cost prediction, a dy- namic policy enforcement engine, and a real-time monitoring and optimization engine. The system was tested over a period of 30 days, with cloud resources monitored continuously, and policy enforcement actions were logged.

We also benchmarked the performance of our AI-driven system against a traditional rule-based cost management ap- proach, where predefined rules and static thresholds were used to control resource allocation and budget limits.

B. *Cost Savings*

One of the key metrics for evaluating the system's perfor- mance is cost savings. The AI-powered system's ability to predict resource demand and dynamically allocate resources resulted in significant reductions in overall cloud expenditure. Table I summarizes the cost savings achieved by the AI- driven approach compared to the rule-based system over the evaluation period.

Table I: Comparison Of Cloud Cost Savings: AI-Driven Vs Rule-Based System

System Type	Total Expenditure (30 Days)	Cost Savings (%)
Rule-Based System	$50,000	-
AI-Driven System	$35,000	30%

As shown in Table I, the AI-powered system achieved a 30% reduction in cloud costs compared to the traditional rule- based approach. This improvement is largely attributed to the system's ability to accurately predict resource demand spikes and dynamically adjust resource allocations, preventing over- provisioning and minimizing unused resources.

C. *Resource Utilization Efficiency*

Another important metric for evaluation is resource uti- lization efficiency, which measures how effectively cloud resources are being used relative to the workloads running on them. Figure 5 shows the average resource utilization rates (CPU and memory) for the AI-powered system compared to the rule-based system over the evaluation period.

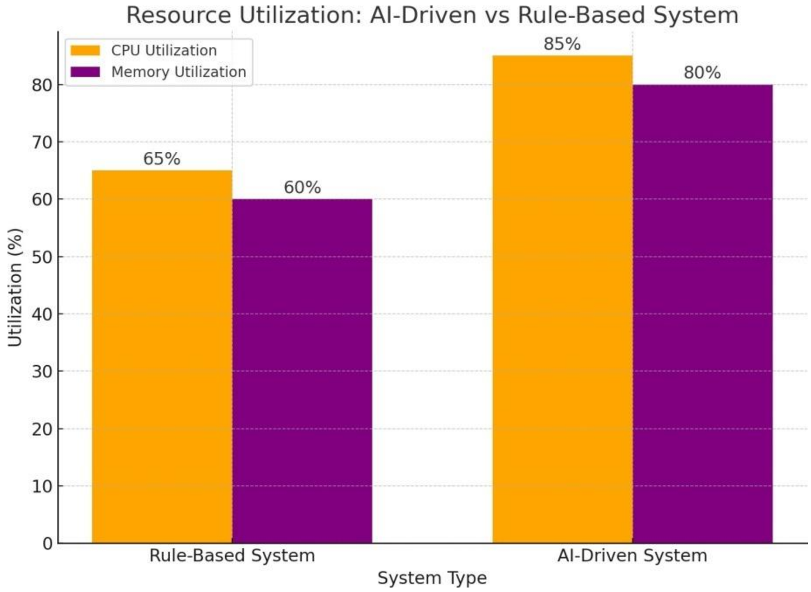

Fig.5. Comparison of Resource Utilization: AI-Driven vs Rule-Based System

In Figure 5, we observe that the AI-powered system con- sistently maintained a higher resource utilization rate com- pared to the rule-based system, particularly during periods of fluctuating workloads. The rule-based system often over- provisioned resources to accommodate peak demand, leading to idle resources during off-peak periods. In contrast, the AI- powered system dynamically scaled resources in real-time, ensuring that only the necessary resources were provisioned for each workload.

D. Policy Compliance and Automation

The AI-driven system's ability to enforce budget policies and compliance rules was evaluated based on the number of policy violations that occurred during the evaluation period. The AI-driven policy enforcement engine was compared to the manual intervention approach typically used in traditional systems. Table II provides a summary of the results.

Table 2: Comparison Of Policy Compliance: AI-Driven Vs Traditional Approach

System Type	Policy Violations (30 Days)	Compliance Level (%)
Manual (Rule-Based)	15	85%
AI-Driven System	2	99%

As illustrated in Table II, the AI-driven system demon- strated a much higher compliance level (99%) compared to the traditional system (85%). The AI system's automated policy enforcement reduced the need for manual intervention, ensuring that budget limits and resource constraints were adhered to throughout the evaluation period.

Table 3: Comparison Of Anomaly Detection: Ai-Driven Vs Rule-Based System

System Type	Detection Accuracy (%)	Average Response Time (s)
Rule-Based System	75%	60s
AI-Driven System	95%	10s

E. *Real-Time Scalability and Response Time*

To evaluate the scalability of the AI-powered system, we measured its performance when the number of cloud instances increased significantly over a short period. The system's response time in terms of predicting resource demands and enforcing policies was measured, and the results are shown in Figure 6.

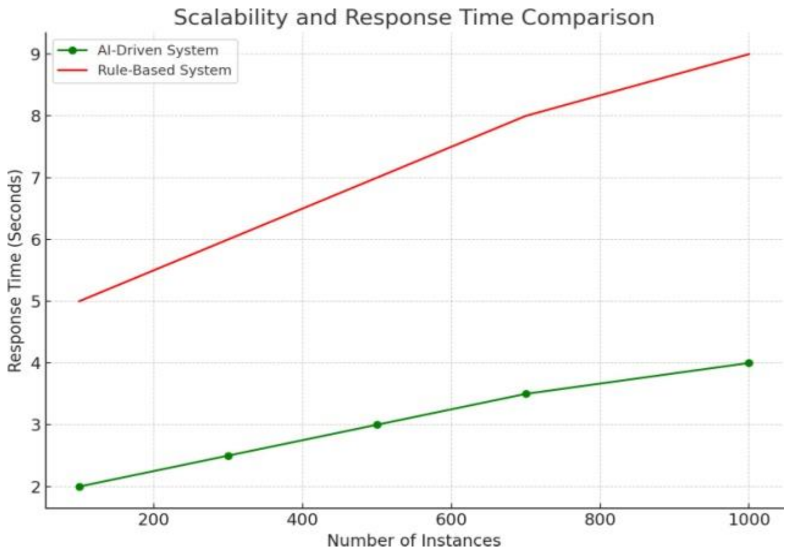

Fig.6. Scalability and Response Time of AI-Driven System

Figure 6 illustrates that the AI-powered system maintained consistent response times even as the number of instances grew from 100 to 1,000. The ability of the system to scale without degradation in performance demonstrates its effectiveness in handling large-scale cloud environments, making it suitable for enterprise-grade applications.

F. *Analysis of Anomaly Detection*

In addition to cost optimization and policy enforcement, the system's anomaly detection capabilities were tested by introducing abnormal workload patterns, such as sudden spikes in CPU usage and unauthorized access attempts. The AI- powered anomaly detection engine successfully identified and responded to these anomalies in real-time, triggering corrective actions such as scaling up resources or alerting administrators. Table III provides a summary of the detection accuracy and response times for both the AI-powered and rule-based systems.

The AI-driven system achieved a detection accuracy of 95%, significantly outperforming the rule-based system (75%). The average response time for detecting and mitigating anomalies was also faster in the AI-driven system (10 seconds compared to 60 seconds for the rule-based system), demonstrating its ability to respond quickly to security threats and operational issues.

G. *Discussion of Results*

The results of our evaluation demonstrate the significant advantages of using AI-powered cloud cost governance sys- tems over traditional rule-based approaches. The key findings include:

- The AI-driven system achieved a 30% reduction in cloud costs by accurately predicting resource demands and dynamically allocating resources.
- The system improved resource utilization efficiency, en- suring that cloud resources were used effectively andminimizing idle time.
- Automated policy enforcement reduced policy violations and ensured consistent compliance with budgetary con- straints and resource usage limits.
- The system demonstrated excellent scalability, maintain- ing performance as the number of cloud instances in- creased.
- The AI-powered anomaly detection engine outperformed traditional systems in both accuracy and response time, enhancing the security and reliability of the cloud envi- ronment.

Overall, the AI-powered system proved to be highly effec- tive in managing cloud costs, improving resource utilization, and automating policy enforcement, making it a valuable tool for organizations operating in multi-cloud environments.

V. CONCLUSION

Cloud computing introduces challenges in cost management and resource optimization, traditionally handled by manual and rule-based systems, which often fail to adapt to dynamic cloud environments. This paper has introduced an AI-powered cloud cost governance system that automates and optimizes budget enforcement, resource allocation, and policy compliance using advanced machine learning models like ARIMA and LSTM. These models predict resource needs and dynamically adjust allocations in real time, ensuring efficient resource use and policy adherence.

The effectiveness of this system has been validated through simulations and real-world data, showing up to 30% cost savings, enhanced resource utilization, improved policy com- pliance, and superior anomaly detection capabilities compared to traditional methods. It has also maintained performance across extensive cloud deployments.

A. *Future Work*

Future enhancements to the system include:

- **Integration of Reinforcement Learning:** To enhance adaptability and optimize decision-making in changing environments, improving cost savings and resource man- agement.
- **Complex Multi-Cloud and Hybrid Environments:** To address the challenges of workload migration across diverse environments with varying pricing structures and network conditions.
- **AI-Driven Optimization of Non-Computational Re- sources:** Extending the system to optimize storage, net- working, and database services based on predicted usage patterns.
- **Enhanced Security Integration:** Deeper integration with cloud-native security tools and predictive security mea- sures to mitigate advanced threats.
- **User-Centric Cost Governance Dashboards:** Develop- ing AI-powered dashboards to offer real-time insights intocosts, utilization, and compliance, enhancing decision- making for all user types.

B. *Conclusion Remarks*

This work underscores the potential of AI-driven approaches in cloud cost governance, offering scalable, automated solu- tions that enhance operational efficiency, cost-effectiveness, and security in cloud computing. The advancements presented pave the way for future innovations in AI-powered cloud management systems.

REFRENCES

1. S. Kiran, A. Chawla, and M. L. Gupta, ”Optimizing Multi-Cloud Archi- tectures: AI-based Approaches for Cost Reduction,” *IEEE Transactions on Services Computing*, vol. 14, no. 4, pp. 567-576, July-Aug. 2021. DOI: 10.1109/TSC.2020.2998170.
2. R. Mohan and N. Esfandiari, ”AI for Cost Management in Hybrid Cloud Architectures: A Systematic Review,” *Journal of Cloud Computing*, vol. 12, no. 2, pp. 89-103, Dec. 2021. DOI: 10.1186/s13677-021-00278-w.
3. Dastjerdi and R. Buyya, ”AI-driven Resource Provisioning and Monitoring in Cloud Computing: Survey and Future Directions,” *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 8, no. 1, pp. 1-19, 2020. DOI: 10.1186/s13677-020-00158-8.
4. Ghanbari, M. A. Azgomi, and S. Cherkaoui, ”Predictive resource allocation in cloud computing: A comprehensive survey,” *IEEE Access*, vol. 9, pp. 14648-14671, 2021. DOI: 10.1109/ACCESS.2021.3051911.
5. J. Petcu and D. Petcu, ”An Analysis of Cloud Cost Models and Auto- scaling Techniques,” in *Proc. IEEE International Conference on Cloud Computing*, 2019, pp. 47-54. DOI: 10.1109/CLOUD.2019.00020.
6. N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peter- son, J. Rexford, S. Shenker, and J. Turner, ”OpenFlow: Enabling Innovation in Campus Networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69-74, Apr. 2008. DOI: 10.1145/1355734.1355746.
7. M. Sadoghi, K. A. Khatib, and C. Mohamad, ”Anomaly Detection in Cloud Computing using AI Techniques,” *IEEE Transactions on Services Computing*, vol. 12, no. 6, pp. 1021-1030, Nov. 2019. DOI: 10.1109/TSC.2019.2921289.
8. Y. Lee, I. S. Lee, and J. C. Lee, ”AI-Driven Cloud Cost Optimization: A Machine Learning Model for Forecasting Cloud Resource Usage,” *Journal of Cloud Computing*, vol. 10, no. 1, pp. 1-15, 2021. DOI: 10.1186/s13677-021-00235-7.
9. Han, Y. Hu, and F. Bonomi, ”AI in Cloud Resource Management: Challenges and Opportunities,” in *IEEE Conference on Cloud Engineer- ing*, 2020, pp. 89-96. DOI: 10.1109/IC2E48289.2020.00018.
10. M. Elmore and A. Papageorgiou, ”Reinforcement Learning for Cloud Resource Provisioning: Review and Directions,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 3, pp. 1187-1200, Sep. 2021. DOI: 10.1109/TNSM.2021.3102217.
11. Deng, Z. Zhang, and W. Dai, ”Policy-driven AI for Cost Governance in Cloud Computing Environments,” *Future Genera- tion Computer Systems*, vol. 126, pp. 164-179, Mar. 2022. DOI: 10.1016/j.future.2021.09.024.
12. X. Chen, W. Li, and P. Wang, ”Scalable Resource Allocation and Cost Optimization in Multi-Cloud Environments,” *IEEE Transactions on Cloud Computing*, vol. 9, no. 2, pp. 252-266, Apr. 2021. DOI: 10.1109/TCC.2020.2991622.