International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| Impact Factor: 6.421|A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103061

# Comparing Major Cloud Providers for AI/ML Workloads: AWS vs Azure vs GCP

**Amruta Gadekar, Nayan Varma, Prerana More, Ashish Salunkhe**

Cloud Engineer, UK

**ABSTRACT:** As artificial intelligence (AI) and machine learning (ML) continue to transform industries, cloud computing platforms have become the backbone for deploying and scaling intelligent applications. Among the leading providers, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have emerged as dominant forces offering a wide range of tools, infrastructure, and services tailored to AI and ML workloads. This paper presents a comparative analysis of these three platforms with a focus on their capabilities, pricing structures, flexibility, ease of use, and performance in handling AI/ML tasks.

The study investigates key service offerings such as AWS SageMaker, Azure Machine Learning, and Google Vertex AI, comparing them across criteria like integration with open-source frameworks, model deployment efficiency, automation of ML workflows, support for custom and pre-trained models, and hybrid/cloud-native compatibility. Furthermore, it assesses the ecosystem and tooling, such as GPU/TPU availability, MLOps features, and data pipeline orchestration.

A combination of qualitative assessment and quantitative benchmarking—through existing performance metrics and service documentation—is used to evaluate the trade-offs each platform presents. Findings suggest that while AWS offers the most mature and extensive suite of services, Azure excels in integration with enterprise systems and DevOps pipelines. Meanwhile, GCP leads in innovative AI services, especially in AutoML and data science tooling.

This paper offers a detailed, side-by-side comparison to assist practitioners, researchers, and enterprises in choosing the most appropriate cloud provider for their AI/ML needs. With the growing complexity and scale of AI deployments, selecting the right cloud platform can significantly impact performance, scalability, and cost-efficiency. The study concludes with insights into how these platforms are evolving and what future improvements may shape the next wave of cloud-based AI/ML development.

## I. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) have become critical drivers of innovation across numerous industries, including healthcare, finance, manufacturing, and e-commerce. These technologies demand scalable infrastructure, access to large datasets, and robust computational resources, which has led to the widespread adoption of cloud computing. The cloud offers on-demand scalability, flexible pricing models, and a rich ecosystem of AI/ML tools, making it the preferred environment for developing and deploying intelligent applications.

Among the major cloud providers, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) dominate the market. Each offers a comprehensive set of services designed to support the full lifecycle of AI/ML workflows—from data ingestion and preprocessing to model training, deployment, monitoring, and governance. However, the choice between these providers is not straightforward. Differences in performance, toolchain compatibility, service maturity, pricing models, and integration options can significantly influence the success of an AI/ML project.

AWS, as the largest cloud provider, has built a mature ecosystem for ML through services like Amazon SageMaker, which provides managed Jupyter notebooks, model training, hyperparameter tuning, deployment, and monitoring. Azure offers Azure Machine Learning, which is deeply integrated with Microsoft's enterprise tools and provides a streamlined pipeline for ML operations (MLOps). GCP, leveraging its AI-first strategy, delivers advanced features through Vertex AI, along with access to custom Tensor Processing Units (TPUs) and AutoML capabilities for rapid model prototyping.

This paper aims to provide a comparative evaluation of AWS, Azure, and GCP for AI/ML workloads. It explores the functional strengths and weaknesses of each platform by examining service features, performance benchmarks, cost-

effectiveness, and developer experience. By highlighting real-world use cases and best practices, the study equips decision-makers with insights to choose the platform that best aligns with their technical and business objectives.
In an increasingly competitive AI landscape, understanding the nuances of each cloud platform is essential for maximizing return on investment, maintaining agility, and accelerating innovation. This research addresses this need through a structured comparison that goes beyond marketing claims to deliver a practitioner-focused perspective.

## II. LITERATURE SURVEY

The growing adoption of AI and ML in industry and academia has catalyzed extensive research into the role of cloud computing in supporting these technologies. A review of recent literature reveals three key themes: the evolution of cloud-native ML platforms, performance and cost comparisons across providers, and the integration of MLOps capabilities.

Several studies have explored the architectures and service models offered by major cloud providers. For instance, Smith et al. (2021) analyze the ML development lifecycle across AWS, Azure, and GCP, emphasizing how managed services like SageMaker, Azure ML, and Vertex AI reduce operational overhead through automation and built-in support for model monitoring, versioning, and governance. Similarly, Khan and Yu (2020) examine the scalability of ML workloads on cloud infrastructure, finding that cloud-native frameworks can achieve near-linear performance improvements with proper resource allocation.

In terms of performance benchmarking, Gupta et al. (2022) conduct experiments using standardized ML models and datasets (e.g., ResNet, BERT) to evaluate training speed, inference latency, and resource utilization across different virtual machines and accelerators (GPUs, TPUs). Their findings highlight that while AWS provides greater flexibility and customization, GCP excels in optimized hardware performance, particularly with TPUs.

Cost analysis is another focal point. According to Li and Chen (2021), pricing models vary significantly between providers, especially when considering long-term workloads, reserved instances, and spot pricing. Azure was found to offer competitive pricing for enterprise users leveraging Microsoft licensing discounts, while GCP's sustained use discounts provide a cost advantage for continuous training operations.

The integration of MLOps practices is also widely discussed. Studies show that Azure ML provides the most mature CI/CD pipeline tools due to its integration with GitHub Actions and Azure DevOps. In contrast, GCP and AWS have been strengthening their offerings with services like Vertex Pipelines and SageMaker Pipelines, respectively.
The literature also notes a trend toward hybrid and multi-cloud deployments, where organizations leverage the unique strengths of multiple platforms. For example, a company might train models using GCP's TPUs and then deploy them using AWS Lambda or Azure Functions for broader integration.

Overall, the literature suggests no one-size-fits-all solution. The choice of cloud provider for AI/ML workloads should be driven by specific use case requirements, existing cloud ecosystems, cost constraints, and team expertise.

## III. RESEARCH METHODOLOGY

This study adopts a qualitative-comparative methodology to evaluate the strengths and weaknesses of AWS, Azure, and GCP for AI/ML workloads. The goal is to provide a practical, side-by-side assessment based on current capabilities, user experience, and strategic alignment with common machine learning use cases.

The research begins with an extensive review of documentation, whitepapers, and service manuals published by Amazon, Microsoft, and Google. Key services analyzed include Amazon SageMaker, Azure Machine Learning, and Google Vertex AI. The comparison focuses on features such as model training and deployment, support for open-source frameworks, automation, MLOps tools, GPU/TPU availability, and pricing strategies.

To complement this theoretical comparison, the study incorporates insights from recent third-party benchmarking studies and real-world case examples sourced from publicly available cloud performance reports and peer-reviewed academic articles. Metrics such as training speed, inference latency, and total cost of ownership (TCO) are included where available.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| Impact Factor: 6.421|A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103061

Additionally, online user forums, technical blogs, and cloud solution engineering reports are analyzed to assess developer sentiment, tooling maturity, and ease of integration with DevOps pipelines. This helps reflect the practical usability of each platform from a practitioner's perspective.

The study does not perform direct hands-on benchmarking due to scope limitations but relies on consolidated findings from credible sources. The platforms are evaluated against a structured rubric that includes criteria across performance, flexibility, ecosystem support, automation, and pricing.

This methodology allows for a comprehensive yet focused comparison, highlighting not just which platform performs better in absolute terms, but which aligns best with particular use case demands—be it prototyping, enterprise-scale deployment, or cost-optimized training.

## IV. KEY FINDINGS

This comparative analysis reveals several important insights into how AWS, Azure, and GCP address the needs of AI/ML workloads. While all three platforms offer comprehensive and evolving AI ecosystems, their unique strengths position them differently for various use cases.

**Amazon Web Services (AWS)** stands out for its maturity, breadth of services, and global infrastructure. Amazon SageMaker offers end-to-end ML capabilities, from data labeling and training to deployment and monitoring. It supports a wide range of frameworks (TensorFlow, PyTorch, MXNet) and provides extensive automation through SageMaker Pipelines and Autopilot. AWS is particularly strong in customization, scalability, and integration with enterprise-grade tools.

**Microsoft Azure** excels in its seamless integration with the broader Microsoft ecosystem, making it highly attractive for organizations already using Azure Active Directory, Office 365, and Azure DevOps. Azure Machine Learning provides strong MLOps capabilities, with easy CI/CD integration and features like ML Designer and AutoML. Azure also offers responsible AI tools such as model interpretability and fairness analysis, positioning it well for regulated industries.

**Google Cloud Platform (GCP)** is widely recognized for its innovative AI features. Vertex AI simplifies the process of building, training, and deploying models, especially with AutoML and pre-trained APIs for vision, language, and translation tasks. GCP leads in specialized hardware with TPUs, which offer excellent performance for deep learning workloads. It is also developer-friendly, thanks to its native integration with TensorFlow and a clean, code-first interface.

Overall, AWS is best suited for highly scalable and complex deployments, Azure is optimal for enterprise and hybrid-cloud environments, and GCP is ideal for rapid experimentation and research-intensive applications. The selection ultimately depends on organizational goals, budget, and technical requirements.

## V. WORKFLOW

The typical AI/ML workflow across cloud platforms consists of several key stages: data ingestion, preprocessing, model training, evaluation, deployment, and monitoring. While AWS, Azure, and GCP follow similar high-level workflows, they differ in how each stage is implemented, automated, and integrated.

### 1. Data Ingestion and Storage

All three platforms provide robust tools for data ingestion from structured, semi-structured, and unstructured sources. AWS offers S3, Azure uses Blob Storage, and GCP utilizes Cloud Storage. These services integrate with data pipeline tools such as AWS Glue, Azure Data Factory, and Google Dataflow for ETL operations.

### 2. Preprocessing and Feature Engineering

Jupyter notebooks are widely supported across SageMaker, Azure ML, and Vertex AI. Preprocessing tools are typically integrated with data lakes, and all platforms offer SDKs and APIs for transforming datasets.

### 3. Model Training and Tuning

AWS SageMaker allows custom containerized training environments or use of built-in algorithms. Azure ML provides

| ISSN: 2394-2975 | www.ijarety.in| Impact Factor: 6.421|A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103061

managed training clusters and hyperparameter tuning through its SDK. GCP's Vertex AI offers AutoML as well as full control via custom training jobs with GPUs/TPUs.

**4. Model Evaluation and Validation**
All platforms provide logging, metrics, and evaluation tools. Azure and GCP emphasize responsible AI practices, offering tools for bias detection and model fairness evaluation.

**5. Deployment and Inference**
Model deployment is streamlined with SageMaker Endpoints, Azure Inference Clusters, and Vertex AI Prediction. Each offers autoscaling, A/B testing, and multi-region support.

**6. Monitoring and Lifecycle Management**
SageMaker Model Monitor, Azure ML Monitor, and Vertex AI Model Monitoring provide real-time insights into model drift, latency, and performance. CI/CD integration is a common feature across all platforms.
Each cloud offers a structured yet flexible pipeline that supports experimentation, reproducibility, and scaling, with different degrees of automation, governance, and customization.

**Advantages and Disadvantages**
Each cloud provider brings unique strengths and trade-offs for AI/ML workloads, making it essential to match platform capabilities with project needs.

**Amazon Web Services (AWS)**
*Advantages:*
- Most mature platform with a wide range of services.
- Extensive support for custom environments and frameworks.
- Deep integration with other AWS services like Lambda, Glue, and Redshift.
- Strong ecosystem and community support.

*Disadvantages:*
- Steeper learning curve due to service complexity.
- Can become costly, especially for large-scale GPU usage.
- Slightly less user-friendly for beginners compared to GCP.

**Microsoft Azure**
*Advantages:*
- Excellent integration with enterprise tools like Active Directory, DevOps, and Office 365.
- Strong MLOps features with Azure Pipelines and GitHub integration.
- Focus on responsible AI with fairness and explainability tools.

*Disadvantages:*
- Less flexible in terms of custom tooling and environment control.
- Interface can feel fragmented between services.
- Slower updates to cutting-edge AI capabilities compared to GCP.

**Google Cloud Platform (GCP)**

*Advantages:*
- Innovative tools like Vertex AI and AutoML.
- TPU support for deep learning acceleration.
- Clean, developer-friendly interface and strong TensorFlow support.

*Disadvantages:*
- Smaller enterprise market share may limit some integrations.
- Fewer prebuilt enterprise governance tools.
- Customer support and regional availability may lag behind AWS and Azure.

| ISSN: 2394-2975 | www.ijarety.in| Impact Factor: 6.421|A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103061

## VI. CONCLUSION

As AI and ML continue to shape the future of technology and business, the role of cloud platforms in enabling scalable, efficient, and robust ML workflows has become increasingly significant. This study examined and compared the offerings of the three leading cloud providers—Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP)—to understand their suitability for AI/ML workloads across multiple dimensions.

AWS stands out for its extensive ecosystem and the maturity of services like SageMaker, offering unparalleled flexibility and scalability. Azure differentiates itself with superior enterprise integration, making it the go-to choice for organizations embedded in the Microsoft ecosystem. GCP, on the other hand, leads in AI innovation, particularly in AutoML, TensorFlow integration, and specialized hardware acceleration using TPUs.

Each platform presents a mix of strengths and trade-offs. While AWS offers powerful customization and a wide array of services, it may be complex for new users. Azure shines in managed workflows and responsible AI, but its AI tooling can sometimes lag behind its competitors. GCP appeals to data scientists and researchers with its simplified and code-first environment but lacks some of the enterprise robustness seen in AWS and Azure.

The choice of cloud provider for AI/ML workloads should be guided by an organization's specific needs—whether they are flexibility, compliance, innovation, or ease of use. There is no universal best; instead, the best choice depends on technical goals, cost constraints, team expertise, and integration requirements.

In a rapidly evolving landscape, cloud providers are constantly enhancing their AI/ML offerings. Understanding their current capabilities and future directions is essential for businesses and researchers to make informed decisions that maximize performance, minimize cost, and support sustainable AI development.

## VII. FUTURE WORK

This comparative study serves as a snapshot of the current capabilities of AWS, Azure, and GCP in 2025, but the cloud and AI/ML landscapes are highly dynamic. Future research should focus on evaluating new and emerging services introduced by these platforms as they continue to evolve to meet the growing complexity of machine learning applications.

One area for further exploration is the integration of multi-cloud and hybrid-cloud strategies. As organizations strive to avoid vendor lock-in and leverage the best features of each platform, evaluating cross-cloud operability and orchestration tools will be critical. Studies could also compare how well AWS, Azure, and GCP handle interoperability through APIs, Kubernetes-based deployments, and unified monitoring systems.

Additionally, a performance benchmarking study with real-time experiments would provide more empirical insights into training and inference latency, hardware utilization, and energy efficiency across platforms. Incorporating synthetic and real-world datasets, such as ImageNet or financial time-series data, could strengthen understanding of how each cloud handles specific ML tasks.

The increasing emphasis on ethical AI and responsible development also warrants further research. Future studies could delve deeper into how each cloud provider integrates fairness, accountability, interpretability, and transparency features into their AI platforms—and how organizations actually adopt them in practice.

Moreover, cost optimization for AI/ML workloads remains a complex issue. Comparative analyses that account for usage patterns, spot pricing, reserved instances, and storage/networking costs over time would offer practical guidance for enterprise cost planning.

Finally, with the rise of generative AI, evaluating the support for LLMs (Large Language Models), diffusion models, and real-time inference pipelines on these platforms will be key. As models become larger and more demanding, assessing cloud readiness for future workloads is a critical avenue for future research.

| ISSN: 2394-2975 | www.ijarety.in| Impact Factor: 6.421|A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 3, May-June 2024 ||

DOI:10.15680/IJARETY.2024.1103061

## REFERENCES

1. Gupta, A., Sharma, M., & Rao, P. (2022). *Benchmarking machine learning performance on cloud infrastructure: AWS, Azure, and GCP*. Journal of Cloud Computing, 11(1), 35–49. https://doi.org/10.1186/s13677-022-00318-7
2. Khan, M., & Yu, H. (2020). *Scalable machine learning with cloud computing: A comparative study*. International Journal of Computer Applications, 975(8887), 22–29. https://doi.org/10.5120/ijca2020909950
3. Li, X., & Chen, Z. (2021). *Cost-performance analysis of cloud-based ML services*. Proceedings of the 2021 IEEE Cloud Conference, 45–52. https://doi.org/10.1109/CLOUD52041.2021.00011
4. O Krishnamurthy. Genetic algorithms, data analytics and it's applications, cybersecurity: verification systems. International Transactions in Artificial Intelligence , volume 7 , p. 1 - 25 Posted: 2023
5. Smith, R., Tan, D., & Zhao, L. (2021). *Machine learning lifecycle management on public cloud platforms*. ACM Computing Surveys, 54(5), 1–30. https://doi.org/10.1145/3442377
6. Dhruvitkumar, V. T. (2021). Autonomous bargaining agents: Redefining cloud service negotiation in hybrid ecosystems.
7. Microsoft. (2024). *Azure Machine Learning documentation*. Microsoft Docs. https://learn.microsoft.com/en-us/azure/machine-learning/
8. Google. (2024). *Vertex AI overview*. Google Cloud Documentation. https://cloud.google.com/vertex-ai/docs/general/overview