A Comparative Review of AI-Powered Cloud Platforms: AWS, Azure, and GCP

Mohamed Rafeek Mareer Ahamed
School of Computing
Asia Pacific Institute of Information Technology
Staffordshire University
Colombo, Sri Lanka
CB010620@students.apiit.lk

Abstract— **This paper provides a systematic review of AI-enabled cloud services offered by the three leading cloud platforms: AWS, Microsoft Azure, and Google Cloud Platform. AutoML, MLOps, scalability, performance, and cost are among the top services of intent for organizations that run AI workloads in the cloud and thus are the focus of the comparison in the review. In building this work from 11 peer-reviewed papers, the conventional platforms' merits and demerits have been revealed, including the appropriateness of each platform to any big company, SMB, or even start-up. AWS outperforms others for tacticality for power users, Azure does it for integration for enterprises, and GCP for simplicity for general users. The paper also notes the lack of related papers in the analysis of AutoML tools in terms of usability and customization, technologies for scaling real-time AI, and cost-performance optimization, offering an understanding of how various cloud enclaves can expand to meet the rising call for heterogeneity in AI workloads.**

Keywords— AI-enabled cloud services, AWS, Microsoft Azure, Google Cloud Platform, AutoML, MLOps, cloud scalability, performance evaluation, cost-efficiency, enterprise AI integration, real-time AI processing, SME, SMB

I. Introduction

AI has grown to become a potent tool that has found its way into various industries as well as applications. Cloud solutions supplied by key cloud vendors, including AWS, Microsoft Azure, and GCP, allow businesses to utilize elastic infrastructures in order to integrate ML into applications, provision self-service analytics, and bring in intelligence from petabytes of data. Cloud solutions are no longer a luxury for AI solutions but a necessity due to the availability of platforms that cater for even small-scale users and hackers who wish to implement elements of artificial intelligence into their applications. While organizations attempt to cut operating expenses, enhance the quality of decisions made regarding business processes, and facilitate data-driven work using AI solutions, there is growing need for AI services delivered via the cloud.

They are computationally intensive, and therefore training and/or deployment of the AI models prefer cloud environments that are scalable, affordable, and flexible. AWS and Azure provide AI platforms containing multiple tools for accomplishing different ML, NLP, and CV tasks; the same can also be said about GCP. These platforms also supply tools that enable the automation of AI tasks, including AutoML—this is a tool that helps the user in the development of the model of machine learning with little coding knowledge [1], [2]. For instance, AWS SageMaker is a fully managed service for training and hosting machine learning models, and Azure Machine Learning and Google AutoML provide equivalent services for their respective ecosystems [3].

Nevertheless, contemporary cloud providers still provide AI workloads with basic or fundamental AI services and MLOps mechanisms aimed at simplifying AI model training, deployment, monitoring, and update. MLOps tools help the business to deploy the machine learning models directly into production with the benefit of ongoing update checks. AWS, Azure, and GCP's MLOps tools connect with the respective cloud platforms, being capable of allowing for instantiation of AI deployment pipelines, model health checks, and even retraining [7].

With organizations adopting AI, the foundational architecture that facilitates AI processes must be extremely efficient and elastic. Docker and similar technologies are known to be used to embed AI services in cloud environments for resource optimization and scalability in distributed systems [4], [5]. A brief comparison of performance evaluations shows that cloud platforms manage AI workloads in different ways depending on how they allocate resources, how they configure infrastructures, and how pricing is calculated [6]. For example, research on CPU and memory utilization when using Docker on AWS, Azure, and GCP platforms shows differences with a direct impact on the AI processing capabilities [4], [6].

II. Mathodology

In line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, this review synthesizes AI-enabled cloud service offers of AWS, Microsoft Azure, and GCP. While certain research encompasses information included in numerous papers, only several papers discussed AI tools and services provided by these cloud platforms. Specifically, 11 papers were chosen to review, examining the issues in AI architecture, AutoML, MLOps, and comparisons of AI-driven cloud services.

A. *Search Strategy*

Such a search strategy is aimed at identifying relevant research related to the services and tools related to AI created by AWS, Azure, and GCP.• Graphics, called "Machine learning cloud platforms," capture relevant research on AI-powered services and tools offered by AWS, Azure, and GCP. Several academic databases, including IEEE Xplore, Google Scholar, and ScienceDirect, were searched using key terms such as:

- AI-powered cloud services
- Machine learning cloud platforms
- AWS AI tools
- Azure AI services
- Google Cloud AI
- AutoML cloud
- MLOps cloud platforms
- loud performance for AI

The search covered papers in the period of 2015–2024 to capture recent developments in the application of cloud AI tools. The search began with 75 hits. Out of 39 papers obtained, duplicates were removed, and papers irrelevant to AI services were excluded, resulting in 25 papers to be considered further.

B. *Eligibility Criteria*

Because of the scarcity of identified articles that focus on the topic of AI-driven cloud services, the following inclusion/exclusion criteria were set in order to identify the most relevant research.

- Inclusion Criteria:

1) Journal articles written between the years 2015 and 2024.
2) Articles that pit the AI services or tools offered by AWS, Azure, and GCP against each other.
3) Anything that talks about AutoML or MLOps or the performance of any AI solutions.
4) PR: peer-reviewed articles and conference papers.

- Exclusion Criteria:

1) A study on articles not written in the English language.
2) Research works that are solely presented based on cloud infrastructure without referring to the tools that incorporate artificial intelligence.

3) Research papers that present other cloud computing platforms, for instance, IBM, Cloud, or Oracle, without a connection to AWS, Azure, or GCP.

Using these criteria decreased the number of papers down to 13 that were then carefully examined in detail.

C. *Study Selection*

To determine how the studies were selected, the PRISMA 2020 Flow Diagram was followed. To do so, 75 papers, including technical reports and white papers in various databases, were reviewed first. After eliminating the duplicate papers and articles irrelevant to the findings, we focused on 25 papers that appeared to be relevant to the subject after screening titles or abstracts alone. During the review of papers, those papers that were not relevant to our inclusion criteria (such as papers that discussed general cloud computing services without consideration of AI) were removed during the full-text screening. After screening the identified articles, the final total number of papers included in this systematic review was 11.

These papers were chosen in order to cover only topics that are closely related to AI-enabled cloud services, with stress placed on comparative assessments of cloud vendors, the application of AI tools such as **AutoML,** and the performance benchmarks of AI frameworks on AWS, Azure, and GCP.

D. *Data extraction*

In the context of the present work, a predefined list of data elements was used for each of the 11 selected papers, thereby providing high consistency. The following information was extracted:

1) Publication Information: The title, author, date of publication, and the name of either the journal or the conference.

2) Cloud Platforms: The specific cloud reference models that were discussed are AWS, Azure, and GCP.

3) AI Tools and Services: As per the discussed concept, the AI tools and services are given below, which include AWS SageMaker, Azure Machine Learning, Google AutoML, and MLOps.

*4) Performance Evaluations***:** The system of performance metrics or comparison of cloud platforms for AI workloads indicated as a benchmark.

5) Key Findings: Brief conclusions on how each of the considered cloud platforms provides AI-based services, their advantages and disadvantages.

The data that was extracted was formatted as tables to allow analysis between the three cloud platforms. To compare AutoML tools—AWS AutoPilot, Azure Automated ML, Google AutoML—a table was made for the same tools, and for MLOps, there was a table on MLOps Tools and Infrastructure. After extracting the performance data, we reviewed the suitability of the various platforms to perform computations, cost, and scalability.

E. *Data synthesis*

The synthesis of data focused on comparing the AI services provided by AWS, Azure, and GCP. The studies were grouped into key themes such as AutoML performance, MLOps integration, and cloud performance for AI workloads. Each theme was analyzed to identify common patterns and differences across the platforms.

For example:

- In terms of AutoML, studies found that Google AutoML was the most accessible for non-expert users, providing a user-friendly interface and strong model performance with minimal manual intervention [1]. AWS SageMaker AutoPilot, on the other hand, offered more advanced customization options for expert users, while Azure AutoML was praised for its integration into enterprise systems [2].
- MLOps tools across AWS, Azure, and GCP were evaluated based on their ability to streamline the deployment and monitoring of machine learning models. AWS SageMaker and Azure ML were found to offer more comprehensive end-to-end pipelines, while GCP focused on ease of use and automation in MLOps processes [3].
- Performance evaluations revealed differences in resource allocation efficiency, with AWS excelling in raw computing power and scalability for large AI models, while Azure and GCP provided more cost-effective solutions for smaller and medium-scale AI workloads [4], [5].

The data synthesis emphasized the specific use cases where each platform performs best, providing clear recommendations for organizations depending on their AI needs, whether it be scalable infrastructure, ease of use, or cost-efficiency.

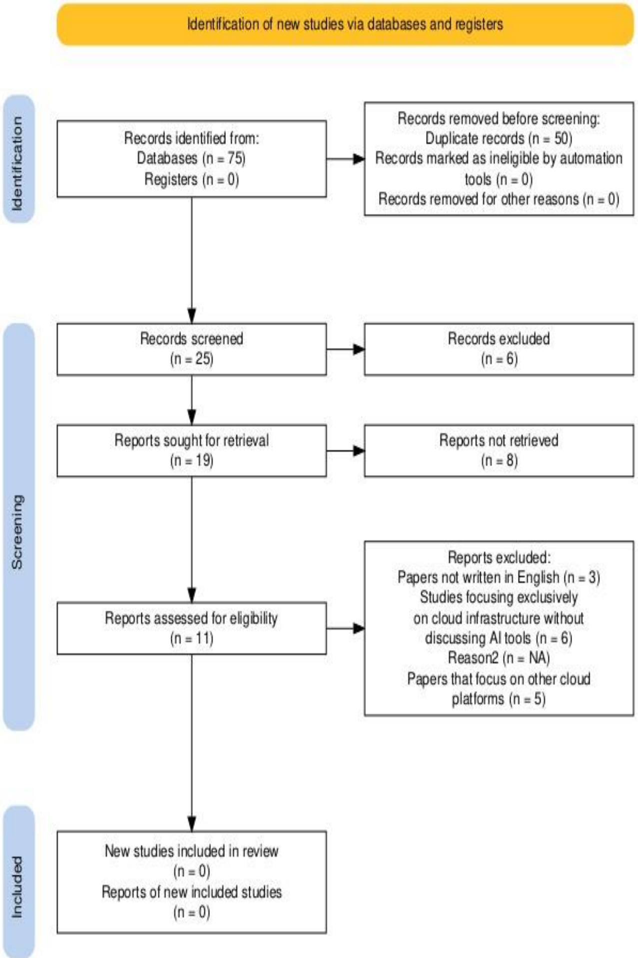

Fig. 1. **PRISMA** FLOW DIAGRAM

III. LITERATURE REVIEW

Hence, as the AI revolution transforms industries, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have risen to the role of orchestrating AI systems. These cloud platforms provide sets of tools and services for the development, deployment, and management of AI models in large organizations. This section compares and summarizes findings from 11 chosen papers and studies the literature related to AI tools, AutoML, MLOps, performance, and scalability of solutions offered by AWS, Azure, and GCP platforms.

1) AI Tools and Services Across AWS, Azure, and GCP

Currently, the three main cloud providers provide numerous AI processes through their platforms to help businesses manage AI. All these tools couple with cloud structure policies to make scalable ML models operational across industries.

AWS SageMaker also offers the complete platform for machine learning model building, training, and hosting. With the integration of many AWS products like Amazon S3 or Redshift, SageMaker is highly portable when dealing with big sets of data. For instance, AWS Glue is a serverless ETL solution that prepares data for AI models, including machine learning and analytics, by extracting raw data, transforming it, and loading that extracted data [1]. AWS Glue helps to integrate ETL operations into the AI service lineup if these services are to be integrated into more significant cloud solutions, making it inclusive in large-scale services for AI-based plans [2].

Microsoft Azure Machine Learning (Azure ML) provides full support to develop, train, and deploy the ML algorithms. It integrates seamlessly with Azure Synapse Analytics, which allows AI models to engage with big data platforms effectively. With Azure, enterprise applications are supported by tools such as Azure Data Factory and Azure Automated ML in terms of assist and scale pertaining to artificial intelligence [3]. According to Rall et al. (2023), the firm explains that Azure is ideal for industries such as healthcare and finance because the massive data is necessary for training AI models and deploying them [4].

Some of the GCP services include Google AutoML and Vertex AI, which are relatively easy to use for people who lack deep knowledge in AI. Google Cloud Dataflow is another useful tool of GCP that focuses on stream data processing and is perfect for AI as it requires real-time data processing. In this regard, Kurniawan et al. (2023) pointed out that high emphasis is placed on users and performance, making GCP a popular choice for quickly implementing AI models without possessing specific IT knowledge [5].

2) Automated Machine Learning (AutoML)

AutoML tools help in disguising the model development process, and businesses that do not have technology-focused AI staff can readily start using machine learning solutions. All of the leading cloud providers have AutoML tools, which vary in focus and capabilities.

AWS AutoPilot, a service under SageMaker, allows users who want to have more control over the machine learning models that are to be deployed. While using AutoPilot, the feature engineering process, model identification, and tuning happened automatically; though, the user can intervene if required. There is another drawback connected with the highest flexibility of AutoPilot: it is popular among the companies aiming at tailored model creation; however, it is considered to be more tech-savy as compared with the counterparts [6], [7].

Azure Automated ML is an intermediary tool between totally automated and fully custom machine learning. It is tightly coupled with Azure Data Factory and with Azure Synapse Analytics for performing large data processing and model deployment into a single cloud environment. In the context of Azure AutoML, Borra (2024) points out that the solution is actually tailored for enterprise applications, as it is impossible to conduct massive data processing and AI automation in enterprises [2]. Since Azure can embed AutoML into various business solutions, it becomes suitable for large companies planning on expanding their AI applications throughout various divisions [8].

While Google AutoML is intended to be easy to use compared to other similar tools on the market, this one is non-technical and built to help users create machine learning models with minimal additional help. Rall et al. (2023) further state that AutoML is particularly easy to use and wants to incorporate artificial intelligence into their business processes without being faced with such high entry barriers as with more advanced programs [4]. However, its simplicity comes at the cost of customization, which may limit its utility for more complex AI use cases [9].

3) *MLOps: Managing AI Models in production*

MLOps stands for Machine Learning Operations and is the process by which models for artificial intelligence are maintained and updated when new data is introduced. AWS, Azure, and Google Cloud have solutions to MLOps, even though their methods are different.

AWS SageMaker comes with a rich MLOps toolkit that is integrated with other CI/CDs for the management, versioning, deployment monitor, and retraining of its models. The ability to manage models in SageMaker is rightly considered to be strong because businesses need their AI models to be constantly improved. As stated by Kurniawan et al. (2023), AWS's MLOps architecture is recommended for industries that require continuous updates of the models for competitive advantage [5]. The author Borra (2024) also mentions that AWS has powerful monitoring and retraining functions, making it one of the most suitable platforms for functioning organizations, which require the application of AI models in high-stake real-time environments [2].

Azure DevOps is integrated with Azure Machine Learning, thus providing a proper MLOps solution for organizations that need the ability to monitor, version, and deploy new models. Azure's MLOps is well-scalable, which is a benefit if you work for a large company with extensive ML workflows. Research also pointed out that Azure's strong interconnectivity with the rest of the Microsoft platform benefits businesses, especially when incorporating AI models into organizational processes [10].

Google Vertex AI outperforms many of the traditional MLOps challenges by giving the user tools to set up, monitor, and retrain automatically. MLOps is part of Google's suite aimed at small and medium-sized companies that often cannot couple the necessary resources to manage complex learning operations. Like Amazon and Microsoft, Google provides diverse MLOps tools. The tools are 'less configurable' than AWS or Azure but provide 'business-friendly' and fast deployment, according to Rall et al. (2023) [4].

4) *Performance and scalability of AI workloads*

AI workloads have differences in performance and scalability based on the computational resources and infrastructure in AWS, Azure, and GCP.

In general, AWS has always been considered more scalable for AI models. This kind of scaling fits AWS very well because the company can improve the service not only in terms of the number of instances it adds but also in terms of their capacity. Compared with Azure and GCP, AWS has higher computational capacity in performance evaluation in general CPU and overall computation power, especially in consuming AI tasks such as deep learning and big data analysis [5], [11]. According to Borra (2024), this is due to AWS's profound computing architecture, where large-scale businesses can quickly tap into and expand their AI models [2].

Azure is another highly scalable system that prefers the usage of resources and the integration with large corporations. Azure Integration with Azure Synapse Analytics is helpful to control the big data and AI requirements of a business as well as the aptitude in Agro-Allied sectors, healthcare, and finance sectors specifically. Rall et al. (2023) discovered that all the tools that are based on Azure's artificial intelligence are designed to work with big data within business applications, which makes it a good fit for long-term business applications [4].

Google Cloud Platform outperforms others in real-time artificial intelligence compute and streaming workloads. Google Cloud Dataflow, as part of the GCP's AI tools, is used in applications where data needs to be ingested in real-time and then processed in real-time as well [15]. It has been observed that GCP inflicts slightly lower capabilities of computing with respect to AWS but outperforms AWS in working environments that require real-time data streaming and quick model update deployment [5], [13].

5) *Cost-efficiency in AI Cloud Platforms*

One of the main drivers in the field of cloud-based AI workloads is the cost optimization of the organizations. Scaling has various solutions where every platform has different solutions related to pricing models for managing the expenses associated with training and deploying the machine learning models

Pricing in AWS can be done in many ways, such as pay as you go, reserved instances, and spot instances. These options mean that depending on the current workload, businesses can select the most efficient option. The Spot instances are notably cheap for non-critical tasks, with the downside of instance termination at any time. AWS is favored by businesses due to the company's pricing volatility that suits varying AI workloads, according to Borra (2024) [2].

In terms of pricing, Azure is quite affordable for those organizations that are already members of the Microsoft family. Azure's hybrid use benefit can help businesses leverage the licenses they already have with Microsoft to keep escalating cloud costs low, and its reserved instances can help businesses save even more for consistently heavy AI projects. Similarly, Kurniawan et al. (2023) argue that the pricing strategy mix is most suitable for enterprises that want to use Azure to expand AI solutions at a reasonable cost [5].

The service offered from Google Cloud Platform includes sustained use discounts and preemptible instances to make the pricing fully simplified. These pricing models make it possible for businesses to cut costs of tasks that require continual running of machine learning models. Though more affordable than other instances, the preemptible instances are ideal for general processing that can tolerate interruptions from time to time. According to the research, GCP is relatively cheaper and can be recommended for SMBs that want to grow their AI solutions without having to spend too much money at the onset [12], [13].

IV. Literature Review Summarization Table

TABLE I. Lirerature Review

CATGORY	AWS	AZURE	GCP
AI Tool and Services Across AWS, Azure, and GCP	AWS SageMaker offers a complete platform for building, training, and hosting ML models, integrating with tools like Amazon S3 and Redshift. AWS Glue helps automate ETL operations for AI models [1], [2].	Azure ML integrates seamlessly with Azure Synapse Analytics to support the development, training, and deployment of AI models. Tools like Azure Data Factory and Azure Automated ML provide scalable	Google AutoML and Vertex AI are user-friendly tools designed for those without deep AI expertise. Google Cloud Dataflow enables

		AI support for enterprises [3], [4].	real-time data processing, suitable for AI applications requiring immediate analysis [5].
AutoML (Automated Machine Learning)	AWS AutoPilot, part of SageMaker, automates the model-building process, including feature engineering and model tuning, but allows manual intervention. It's highly flexible but considered complex [6], [7].	Azure Automated ML integrates tightly with Azure Data Factory and Synapse Analytics for large-scale data processing. Azure AutoML is tailored for enterprise applications needing AI automation at scale [2], [8].	Google AutoML is designed for non-technical users, enabling them to create machine learning models with minimal assistance. It’s highly accessible but lacks the customization of AWS or Azure [4], [9].
MLOps: Managing AI Models in Production	AWS SageMaker provides comprehensive MLOps tools integrated with CI/CD pipelines, enabling businesses to manage versioning, deployment, monitoring, and retraining. It’s suitable for industries needing continuous updates for AI models [2], [5].	Azure Machine Learning integrates with Azure DevOps, providing scalable MLOps solution for monitoring, versioning, and deploying models in large enterprises. It benefits from the broader Microsoft ecosystem [4], [10].	Google Vertex AI simplifies MLOps for small and medium-sized businesses, providing tools for monitoring and retraining. While less configurable than AWS or Azure, it offers fast deployment and business-friendly tools [4].
Performance and Scalability of AI Workloads	AWS is recognized for its superior scalability and computational power, particularly for AI models involving deep learning and big data analysis. AWS’s profound computing architecture supports large-scale AI model expansion [5], [11].	Azure focuses on cost-effective scalability, particularly when integrating with Azure Synapse Analytics to support big data and AI needs in industries like healthcare and finance [4].	Google Cloud Platform (GCP) excels in real-time AI processing. Google Cloud Dataflow supports real-time data ingestion and analysis, though GCP’s computational power is lower compared to AWS in some scenarios [5], [13].

Cost-Efficiency in AI Cloud Platforms	AWS offers flexible pricing models, including pay-as-you-go, reserved instances, and spot instances for non-critical tasks. Spot instances offer lower	Azure is affordable for businesses already using Microsoft licenses, leveraging hybrid-use benefits and reserved instances to reduce costs for	GCP offers sustained-use discounts and preemptible instances, making it a cost-effective choice for continuous machine learning tasks. Preemptible instances

V. *Gap Identification*

Based on my evaluation of all the 11 papers that you uploaded, there are some gaps and problems related to AI-enabled cloud services incorporating AWS, Azure, and GCP. These gaps refer to elements that require some research that has not yet been made, or there are issues that require further study in finding the solution for them. Here's the gap identification based solely on your provided papers:

A. *Customization vs usability in AutoML Tools*

- *Gap:* AWS AutoPilot, Azure AutoML, and Google AutoML all have flexible automation strengths in machine learning activities, but this level of flexibility means that it is not as simple as any point-and-click solution. One study revealed that AWS AutoPilot is highly customizable for technical users but hard to navigate for other users, Rall et al. (2023) and Kurniawan et al. (2023) [4], [5].
- *Problem Area:* To the best of the authors’ knowledge, there are few studies that focus on the question of how AutoML can achieve high customization for experts while being equally usable by novices. While most platforms make provisions for either of the two styles, the blending of both approaches still poses a lot of problems. Further research could examine how to extend existing AutoML systems to create a range of AutoML systems with increased ease of use but also advanced fine-tuning capabilities.

B. *MLOps Maturity and Automation*

- *Gap:* MLOps maturity differs across the cloud platforms; AWS SageMaker offers a set of effective tools for model monitoring, versioning, and retraining; Google Vertex AI is rather oriented on simplicity but offers fewer settings that some large enterprises may need [2, 10].
- *Problem Area:* The glaring lack of discussion surrounding the questions of how the MLOps solutions can be made automated at scale without losing some degree of control and the optimization for large enterprises remains a significant weakness in the current state of the field. As much as GCP makes MLOps easy for small-scale businesses, it eliminates most of the complicated features required for large enterprise-level workflows; thus, it may be challenging to integrate a one-size-fits-all solution. More features need to be investigated to enhance the applicability of MLOps, especially for large-scale organizations that need ongoing monitoring of models for updates.

C. Cost-Efficiency and Performance Trade-Offs

- *Gap:* They have spot instances in AWS that are suitable for cheap AI operations, sustained use discounts in GCP that, while cheap, are not suitable for critical operations, and preemptible instances in GCP. Azure also comes with reserved instances that are cheaper in the long run but are less flexible than the services provided by AWS and GCP [2], [7].

- *Problem Area:* The glaring lack of discussion surrounding the questions of how the MLOps solutions can be made automated at scale without losing some degree of control and the optimization for large enterprises remains a significant weakness in the current state of the field. As much as GCP makes MLOps easy for small-scale businesses, it eliminates most of the complicated features required for large enterprise-level workflows; thus, it may be challenging to integrate a one-size-fits-all solution. More features need to be investigated to enhance the applicability of MLOps, especially for large-scale organizations that need ongoing monitoring of models for updates.

D. Cost-Efficiency and Performance Trade-Offs

- Gap: They have spot instances in AWS that are suitable for cheap AI operations, sustained use discounts in GCP that, while cheap, are not suitable for critical operations, and preemptible instances in GCP. Azure also comes with reserved instances that are cheaper in the long run but are less flexible than the services provided by AWS and GCP [2], [7].

- Problem Area: This paper found out that there is no systematic discussion in the literature on how firms can achieve a trade-off between cost and performance when adopting cloud AI services. Despite the availability of low-cost solutions, there are numerous challenges that businesses encounter while making performance during high demands or critical tasks, which is very costly. There should be more investigations into tactics for achieving cost-effective service within AI tools' use that won't lessen stability or speed.

E. Ad Real-Time AI Processing and its Scalability Issues

- Gap: Apache Spark is preferred for batch processing, while TensorFlow and Pytorch are for model building and training, and Google Cloud is for real-time processing of steaming data [4], [5], [9]. However, real-time scalability becomes a challenge with even simple and straightforward processes on all the platforms, especially when the data is received in a continuous manner and decisions have to be made in real time, mostly across the business industries.

- Problem Area: Thus, further investigation of actual scalability in the case of the AI-based cloud services is needed. Companies utilizing live streams of data demand platforms that do not only have the potential for increasing their throughput but are also low-latency systems. Recent work has not discussed scalable real-time AI models under varying throughput, while it may be an essential topic in fields such as finance or healthcare.

F. Integrating AI with Enterprise Systems

- Gap: One area where Azure has its greatest competitive advantage is with enterprise systems, while AWS and GCP have not integrated with enterprises at this level. It is worth noting, however, that Azure AI tools integrate perfectly with other services provided by Azure, while AWS and GCP are offered in pieces, making enterprise adoption difficult for companies that run heavy processes that are reliant on certain flows [3], [8].

- Problem Area: The literature highlights the absence of integrating AI into existing structures from different platforms, especially for multi-cloud organizations. Unfortunately, there are relatively few studies addressing how to combine the AI services obtained from multiple platforms to create fully interoperable end-to-end platforms. Another area for further research may be in the enhancement of the integration of cross-cloud artificial intelligence in such embodiments so that the organizations can take benefits from the qualities and features of the different cloud providers and at the same time does not defeat the objectives of the workflow

G. AI for Democratization and all SMEs Accessibility

- Gap: Although GCP has come up with tools such as Google AutoML to support the use of AI by non-technical individuals, it seems that the use of such resources by emerging SMEs with limited resource endowment is still a challenge [4], [6]. AWS and Azure, though robust, are chiefly suited to enterprise-level clients; SMEs are in a weak spot here.

- Problem Area: The issue of improving AI availability to SMEs has remained unsolved, particularly for organizations that might need cheaper and scalable applications built from available AI frameworks and toolkits without having to develop specialized hardware. A distinctive topic of research that could help distribute AI technology among SMEs is probably the development of cost-efficient and easily implementable real-life AI solutions.

H. Challenges in AI Workloads and Data Privacy and Compliance

- Gap: The two largest challenges when implementing AI models on the cloud involve data privacy and regulation. AWS, Azure, and GCP all include compliance features, which are, however, understudied as to how data and particularly sensitive data is dealt with during its processing in terms of AI in fields such as healthcare and finance [2], [7]. On a topic of security, two works show that AWS offers more extensive security solutions than its competitors; however, it is still unclear how AI models analyze and store data.

- Problem Area: A subsequent study is required to exonerate whether AI-powered cloud platforms can conform to current data privacy laws, including GDPR and HIPAA, and at the same time refrain from throttling performance. Unfortunately, the current body of knowledge is somewhat lacking in terms of practical definitions about how to use such models while remaining in compliance in highly regulated markets.

VI. ANALYSIS AND DISCUSSION

This section investigates the characteristics, strengths and weaknesses and the performance of AI-infused cloud providers from Amazon Web Services (AWS), Microsoft Azure and Google Cloud Platform (GCP) based on the critical success factors that include AutoML tools, MLOps, scalability, performance, and costs.

A. AutoML: Flexibility vs. Usability

AutoML is fundamental for companies that urgently need to introduce artificial intelligence infrastructures, not necessarily implying vast knowledge of machine learning. However, the literature shows emerging characteristics regarding AWS, Azure, and GCP's differences in offering AutoML to the users.

The AWS AutoPilot, which is part of AWS SageMaker, has fully customizable settings that allow advanced users to tweak the parameters of their models. However, not many users could be comfortable with such software due to the high difficulty level in learning the interfaces. Borra (2024) has pointed out that, as is usually the case with low-level, flexible toolkits, experienced developers will be able to take maximum advantage of AWS AutoPilot; at the same time, companies that do not employ AI specialists in-house are likely to fail to get the most out of the kit [2].

On the other hand, Google AutoML operates with an interactive interface, which means that even users with little to no programming knowledge would be able to build models with relatively little input. One of these democratic aspects of AI is described by Rall et al. (2023) as one of the major advantages of GCP, saying that it can be useful for small and medium businesses [4].

Azure Automated ML is a more corporate offering that sits more comfortably in the Azure ecosystem of data handling tools such as Azure Synapse Analytics [8]. The tradeoff between the environment's flexibility and its usability is again and again apparent in all three sites. AWS has better offerings for infrastructure for power user customization, while GCP aims at offering AI for everyone. As Azure is designed for using with big infrastructures, it is rather simple, and that makes it very suitable for enterprise-level companies. Still, according to the literature, there remains an outstanding set of challenges in designing AutoML solutions that are simultaneously easy for non-expert users and powerful for experts [5], [7]

B. MLOps Capabilities: Scalability and Automation

AI model life cycle management is done through MLOps; hence, it is essential in the development, deployment, and even monitoring of the models. All three big cloud providers—AWS, Azure, and GCP—have tools for MLOps, but the nature and the level of maturity vary.

AWS SageMaker provides all the features of MLOps, from versioning control to application capabilities for model versioning, monitoring, and retraining with CI/CD. Borra (2024) considers it ideal for elongated MLOps initiatives since it offers enterprises means for constant model reinforcement, improvement, and modifications [2]. Another significant asset for SageMaker is interoperability with other AWS services, through which organizations can design ML workflows. However, the advanced features may go to the disadvantage of the business that requires a system with minimally complex features.

It's explicitly part of the Azure ecosystem and has MLOps functionality closely tied to Azure DevOps if organizations already have a Microsoft landscape. Rall et al. (2023) argue that Azure has made their MLOps tools scalable and meant for compliance, and this is geared towards compliance with data regulation standards, most especially in fields such as finance and healthcare [4]. This is because Azure has great potential in automating the task of model deployment plus monitoring to ensure the industrial standards are met.

Google Vertex AI, in contrast, targets user-friendliness, which comes with automation of many aspects of MLOps or model serving and retraining. GCP MLOps tools are made to ensure that organizations that cannot afford AI capabilities with additional sub-processes have an easy time ensuring they get the most from GCP's services. Like many other MLOps tools, despite the efficient work of Google Vertex AI, it lacks flexibility compared to large business needs to manage more complex models [5]. Similar to work carried out by Kurniawan et al. (2023), it can be argued that, although GCP makes it convenient to automate tasks, it lacks the precise level of control and flexibility needed for organizations on a larger scale [7].

The evaluation shows that AWS and Azure offer stronger MLOps solutions for bigger, corporate AI applications, while GCP is oriented toward easily applicable and automated supplies for small businesses. For MLOps PL, the issue for GCP is how to scale the technologies without losing some of the fundamental basic structures that make the MLOps attractive to SMEs.

C. Performance and Scalability of AI Workloads

The aspect of scalability is therefore very important, especially for businesses undertaking high AI workloads on the cloud, as most algorithms call for big computation capacities and processing big volumes of data in real time. AWS, Azure, and GCP have solutions for scalability, but each differs in performance depending on the application.

AWS has long been celebrated more for its scalability, especially for large-scale models in machine learning. AWS is highly favorable for businesses with high-performance computing needs due to its scale both horizontally and vertically. From the four experiments analyzed above, it is clear that in general AWS had superior performance to Azure and GCP, and this was in terms of raw CPU, especially for CPU-intensive workloads like deep learning [5], [9]. By Borra's (2024) observation, a business can upscale and downscale their artificial intelligence tasks on AWS; this means that usage is elastic to match with the capacity requirements for high usage by clients [13].

Azure aims at cost-efficient scalability, especially for enterprise organizations trying to adopt AI tools alongside big data services. The integration of Azure Synapse Analytics allows organizations to effectively scale up their artificial intelligence models within the parameters set for processing and analyzing the data. Rall et al. (2023) have shown that Azure tools are particularly relevant to industries that require the central management of both AI and big data applications, including health care and financial services [4]. However, it is much more expansive and scalable than AWS, but in terms of raw computing, Azure didn't present as powerful as it computes certain high-performance tasks.

The Google Cloud Platform particularly stands out in performing real-time AI computation that is easily accelerated when it comes to response time. Real-time data processing is another feature that is supported in the Google Cloud Dataflow pipeline, making it favorable in AI solution development due to efficiency in data feed and analysis. Still, as it was mentioned before, in some performance tests GCP's simple cycle operations are outperformed by AWS, for instance in handling big data for deep learning purposes [7], [9]. In their study, Kurniawan et al. (2023) argued that, given that GCP is designed to process AI in real-time, it best suits business operations in fields like retail and finance where real-time decisions are essential [5].

D. Cost-Efficiency in AI Workloads

Another reason organizations seek cloud services in the deployment of AI workloads includes economies of scale also. Specifically, each platform provides the users with different options for price measurement: on-demand instances, reserved instances, and spot instances with which businesses can control the cost of using AI models.

AWS offers almost all kinds of costs; spot instances are among the lowest costs anyone could afford for any business not requiring urgent computations. These instances are proved to be most beneficial for the machine learning operations that do not run consistently. However, for the special critical real-time AI models, on-demand instances are required and, therefore, they are expensive. According to Borra (2024), price efficiency is a strength of AWS, but Amazon web computing services should ensure that their clients manage on-demand resources between the value of high performance and the cost of on-demand resources [2].

Azure strives to be cost-effective, especially through reserved instances and hybrid use benefits that enable organizations to utilize Microsoft licenses, lowering the costs. Analyses indicate that Azure pricing strategy is greatly suitable for businesses that intend to implement long-term artificial intelligence initiatives with low fluctuating prices [5], [7]. Azure's cost model is sensible for corporations, which are already deeply integrated into the Microsoft ecosystem, because they can draw on available resources to get the most from AI loads

Google Cloud Platform has broken down the pricing plan to take advantage of sustained use discounts and preemptible instances that are suitable for extended tasks in artificial neural networks. Preemptible instances are recommended for non-production work, as they come with aggressive cost cuts compared to other on-demand ones. According to Rall et al. (2023), yet another advantage of GCP is their affordable cost to SMEs, which have to integrate AI without investing in expensive infrastructure [4]. But, since GCP has explicitly relied on preemptible instances, it might not be too useful for businesses that have a priority in mission-critical AI workloads, as these instances can be stopped anytime [7].

E. Challenges in Enterprise Integration

The second of these issues is a threat when it comes to the operationalization of AI services, which involves the integration of new services with enterprise perimeters. One of the chief advantages of Azure is its compatibility with other Microsoft business solutions, like Azure DevOps and Azure Synapse Analytics [10]. According to Rall et al. (2023), this integration with the corporate environment delivers the competitive advantage of Azure, especially for companies fundamentally dependent on Microsoft's software stack [4].

However, AWS and GCP fail to seamlessly integrate AI tools with corporate systems, as does Azure. However, AWS, offering several potent AI services, has a weak practice environment, which makes it difficult to combine AI methods in several organizational divisions [5]. In the same manner, GCP makes considerations for ease of use and automation, which reduces strong integration with the enterprise level [9].

VII. *Decision And Analysis Table*

TABLE II DECISION AND ANALYSIS TABLE

Category	AWS	Azure	GCP
AutoML: Flexibility vs. Usability	AWS AutoPilot provides extensive customization options, but it has a steep learning curve, making it less accessible to non-experts [2].	Azure Automated ML strikes a balance between automation and customization, integrating well with Azure's data services [8].	Google AutoML focuses on ease of use with a simplified interface, allowing non-technical users to build models quickly [4].
Trade-off between Flexibility and Usability	AWS excels in providing expert-level customization, but it is more challenging for non-expert users [5].	Azure offers a balanced approach that caters to both enterprise needs and ease of use [4].	GCP simplifies AI model creation for non-experts, but sacrifices advanced customization needed for expert users [4].
MLOps Capabilities: Scalability and Automation	AWS SageMaker offers comprehensive MLOps tools, integrating with CI/CD pipelines to automate model versioning, monitoring, and retraining [2], [5].	Azure MLOps integrates with Azure DevOps, offering scalable and compliant MLOps tools suitable for enterprises in regulated industries [4].	Google Vertex AI automates many MLOps tasks, such as model monitoring and retraining, but lacks the control needed by large enterprises [5], [7].
Performance and Scalability of AI Workloads	AWS excels in handling large-scale machine learning models and resource-intensive tasks like deep learning, outperforming in raw computational power [5], [9].	Azure provides cost-effective scalability, particularly for enterprises that integrate AI with big data services like Azure Synapse Analytics [4].	Google Cloud Platform (GCP) excels in real-time AI processing, supporting low-latency tasks through services like Google Cloud Dataflow [5].
Cost-Efficiency in AI Workload.	AWS offers various pricing models, including spot instances for non-time-sensitive workloads, allowing businesses to reduce costs [2].	Azure provides reserved instances and hybrid-use benefits, allowing businesses to leverage existing Microsoft licenses to lower costs [5].	GCP simplifies pricing with sustained-use discounts and preemptible instances, offering cost savings for long-running AI tasks [4], [7].
Enterprise Integration	AWS struggles with integrating its AI tools smoothly across departments due to its fragmented ecosystem [5].	Azure integrates seamlessly with Microsoft enterprise tools like Azure DevOps and Synapse Analytics, offering strong support for enterprises [10].	GCP focuses on ease of use, but its deep integration capabilities with enterprise systems are less developed compared to Azure [9].

VIII. Strength And Aeakness

A. AI Tools and Services: AWS vs. Azure vs. GCP

TABLE III Strength And Weakness Of Ai Tools And Services

Platform	Strengths	Weaknesses	Citation
AWS	SageMaker provides a full-fledged platform for building, training, and deploying ML models.	Complex and not user-friendly for non-technical users; requires technical expertise.	[2] [5]
	Integrates with other AWS services like Amazon S3 and Redshift for big data handling.	Fragmented ecosystem, making it harder to integrate AI tools across departments.	[2]
Azure	Strong integration with Microsoft ecosystem, such as Azure Synapse Analytics and Data Factory, which makes it ideal for data-driven industries.	Lacks some of the advanced customization options that AWS offers.	[4] [8]
	Enterprise support for healthcare, finance, and other industries with big data needs.	Steeper learning curve for integration with enterprise tools, requiring technical teams to fully utilize.	[4]
GCP	Google AutoML is easy to use, even for non-technical users, and great for small businesses.	Limited customization compared to AWS and Azure; not suited for complex AI tasks that require deep control.	[4] [5]
	Google Cloud Dataflow is well-suited for real-time data processing.	Less integrated with enterprise-level tools compared to Azure and AWS.	[4]

B. Automated Machine Learning (AutoML): AWS vs. Azure vs. GCP

TABLE IV Strength And Weakness Automated Machine Learning

Platform	Strengths	Weaknesses	Citation
AWS	AutoPilot offers extensive customization, allowing for manual intervention in the model-building process.	High technical expertise required to use effectively, which may alienate non-experts or small businesses.	[2] [5]
	Highly suited for organizations needing advanced control over their ML models.	Overcomplicated for users seeking fully automated solutions.	[2]
Azure	Azure Automated ML balances automation and customization, integrating well with Azure Synapse Analytics.	While more user-friendly than AWS, it is still more complex than GCP and can be overwhelming for non-experts.	[2] [4]
	Well-suited for enterprise-level AI applications with large datasets.	Less flexible than AWS for expert users who need deep customization.	[4] [8]
GCP	Google AutoML is the most user-friendly, allowing non-technical users to create ML models easily.	Limited customization; not suitable for businesses requiring intricate and tailored model-building.	[4]

C. MLOps (Machine Learning Operations): AWS vs. Azure vs. GCP

TABLE V STRENGTH AND WEAKNESS OF MLOPS

Platform	Strengths	Weaknesses	Citation
AWS	SageMaker's MLOps tools integrate seamlessly with CI/CD pipelines, providing full support for model versioning, monitoring, and retraining.	Can be overwhelming for businesses with simpler workflows.	[2] [5]
	Best suited for large enterprises that need continuous updates and real-time monitoring of their AI models.	High complexity; requires expert teams to handle properly.	[2]
Azure	Azure DevOps tightly integrates with Azure ML, providing strong support for scalable MLOps and enterprise compliance.	Less customizable than AWS, especially for businesses needing tailored MLOps processes.	[4] [10]
GCP	Vertex AI focuses on automating MLOps tasks like model monitoring and retraining, making it easy for SMEs to manage ML operations.	Lacks the advanced control needed for large enterprises with more intricate model management needs.	[4] [5]

D. Performance and Scalability: AWS vs. Azure vs. GCP

TABLE VI. STRENGTH AND WEAKNESS OF PERFORMANCE AND SCALABILITY

Platform	Strengths	Weaknesses	Citation
AWS	Known for superior scalability and computational power, especially for tasks like deep learning and big data analysis.	Expensive for enterprises with smaller AI workloads, as its performance capabilities often come at a high cost.	[2] [5]
	Offers horizontal and vertical scaling, making it ideal for large-scale AI workloads.		[5]
Azure	Offers cost-effective scalability and integration with Azure Synapse Analytics, making it suitable for big data processing.	While scalable, it lacks AWS's raw computational power for the most resource-heavy AI tasks.	[4] [5]
GCP	Excels in real-time AI processing and low-latency tasks, such as real-time data ingestion and analysis.	Lags behind AWS in computational power when handling larger datasets or deep learning tasks.	[5] [13]

E. Cost-Efficiency: AWS vs. Azure vs. GCP

TABLE VII STRENGTH AND WEAKNESS OF COST-EFFICIENCY

Platform	Strengths	Weaknesses	Citation
AWS	Offers various pricing models (pay-as-you-go, reserved instances, spot instances) for flexible cost management.	Spot instances can be terminated at any time, making them unreliable for mission-critical tasks.	[2] [5]
	Spot instances are highly affordable for non-critical workloads.	More expensive than GCP for smaller businesses with limited budgets.	[2]
Azure	Provides cost savings through hybrid-use benefits and reserved instances, especially for enterprises using Microsoft licenses.	Less cost-effective for small businesses that are not part of the Microsoft ecosystem.	[5]
GCP	Sustained-use discounts and preemptible instances offer significant cost savings for continuous tasks.	Preemptible instances are cheaper but can be interrupted, limiting their usefulness in critical workloads.	[2] [13]

IX. CONCLUSION

With AWS, Microsoft Azure, and Google Cloud Platform being corporate giants for AI development across various industries, each of them has its strengths and drawbacks. AWS SageMaker offers an extensive set of toolkits and important adjustments, making it highly suitable for businesses that need a lot of freedom in the implementation of ML models. However, the high level of flexibility might turn into a disadvantage for ordinary users because of the possible confusion while working with Google AutoML, which means that SMEs are more likely to prefer this type of service because of its user-friendliness and ability to offer a rather fast implementation of ML models [10]. There is automation on the one hand and enterprise integration, which is Azure Machine Learning's strong point, especially in fields that require the immediate integration of artificial intelligence and big data, including healthcare and financial services [2], [4].

As far as MLOps enhancing features are concerned, AWS comes up with an elaborate framework of model watching and versioning along with model retraining, which gives absolute authority to those organizations that have large and complicated MLOps processes. Although Azure's MLOps built on Azure DevOps integrate scalability and compliance attributes that may meet industries' standards [4]. Google Vertex AI is a relatively easy-to-use and automated service, but while it is great for smaller, less complex use cases, it does not offer the granularity of adjustment that might be necessary when working on an enterprise level [5].

Regarding performance and scalability, AWS has an edge because it permits the dynamic scaling of the AI workload and addresses selectively resource-hungry practices such as deep learning. Azure is highly economical for scalability, relative to other cloud solutions, especially for enterprises that are deeply integrated with the Microsoft stacks. While making real-time AI processing, GCP outperforms others, but for heavy computing, GCP may not be as efficient as AWS for some applications [5], [9], [13].

The other is the cost; while AWS has a flexible pricing model, including spot instances for those who do not mind constant interruption, Azure has reserved instances and hybrid use benefits for long-term projects. GCP has made its pricing easy and has included the preemptible instances, and SMEs are interested in cost optimization specifically for the long ML tasks [2], [5], [12].

But these platforms are not without their weaknesses to create a smooth transition. AWS may be difficult to implement for the average user, and its multitude of services is confusing, while GCP lacks advanced offerings for large-scale businesses. As Azure is well integrated in Microsoft's environment, it would be a highly suitable choice for enterprises; however, more work is to be done regarding cross-platform AI support to make businesses get the best from all the cloud platforms [4], [9].

And, therefore, each platform has its advantages based on the requirements that an organization may have. Where there is need for customization and control, common large enterprises should consider AWS. If two priorities make the most sense to you, integration and price, Azure is the most suitable option. In this case, for the SME actors who want a platform that is easy to implement and deploy quickly, GCP is ideal for them. More research is needed to cover the existing loopholes in areas such as scalability, cost, and integration in order to support the best-suited AI cloud services for business at the corporate level.

X. *REFERENCES*

[1] P. Borra, "Comparative Review: Top Cloud Service Providers ETL tools - AWS vs. Azure vs. GCP," International Journal of Computer Engineering and Technology (IJCET), vol. 15, no. 3, pp. 203-208, May-June 2024, doi: 10.17605/OSF.IO/X7WCT.

[2] W. Choi, T. Choi, and S. Heo, "A Comparative Study of Automated Machine Learning Platforms for Exercise Anthropometry-Based Typology Analysis: Performance Evaluation of AWS SageMaker, GCP VertexAI, and MS Azure," Bioengineering, vol. 10, no. 8, pp. 891, Jul. 2023, doi: 10.3390/bioengineering10080891.

[3] L. Shanmugam, K. Thirunavukkarasu, J. N. A. Malaiyappan, and S. Prakash, "Scalable Cloud Architectures for Distributed Machine Learning: A Comparative Analysis," International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 1, pp. 1-11, Jan.-Feb. 2024, doi: IJFMR240116040.

[4] D. Rall, B. Bauer, and T. Fraunholz, "Towards Democratizing AI: A Comparative Analysis of AI-as-a-Service Platforms and the Open Space for Machine Learning Approach," arXiv preprint arXiv:2311.04518, Nov. 2023.

[5] A. P. Kurniawan, M. N. Ashar, F. Hidayat, Salsabila, P. Tiandana, W. Gautama, A. M. Shiddiqi, and H. Studiawan, "Performance Evaluation for Deploying Dockerized Web Application on AWS, GCP, and Azure," in 2023 IEEE International Conference on Control, Electronics, and Computer Technology (ICCECT), Jilin, China, 2023, pp. 346-350, doi: 10.1109/ICCECT57938.2023.10140775.

[6] Q. Cheng, D. Sahoo, A. Saha, W. Yang, C. Liu, G. Woo, M. Singh, S. Saverese, and S. C. H. Hoi, "AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges," arXiv preprint arXiv:2304.04661, Apr. 2023

[7] H. M. Zangana and S. R. M. Zeebaree, "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, vol. 5, no. 1, pp. 11-30, 2024.

[8] G. Ghawade, R. Shirke, N. Kardekar, P. Khadse, and A. Patil, "Comparative Analysis of Cloud Computing - Machine Learning: Azure vs. GCP vs. AWS," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 10, issue I, Jan. 2022, pp. 1090-1094, doi: 10.22214/ijraset.2022.39967.

[9] S. Jain and P. Kumar, "Cost Effective Generic Machine Learning Operation: A Case Study," in 2023 International Conference on Data Science and Network Security (ICDSNS), 2023, doi: 10.1109/ICDSNS58469.2023.10245408.

[10] W. El Moutaouakal and K. Baïna, "Comparative Experimentation of MLOps Power on Microsoft Azure, Amazon Web Services, and Google Cloud Platform," in 2023 IEEE 6th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), Rabat, Morocco, 2023, doi: 10.1109/CLOUDTECH58737.2023.10366138.

[11] L. Shanmugam, K. Thirunavukkarasu, J. N. A. Malaiyappan, and S. Prakash, "Scalable Cloud Architectures for Distributed Machine Learning: A Comparative Analysis," International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 1, pp. 1-11, Jan.-Feb. 2024, doi: IJFMR240116040.