International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing AI through Cloud Infrastructure Empowerment

Ramasankar Molleti

Submitted: 27/01/2024 **Revised**: 10/03/2024 **Accepted**: 19/03/2024

Abstract: The combination of artificial intelligence (AI) with cloud computing implies a promising future of novel technological solutions that have shifted the ways of creating, implementing, and maintaining AI systems. This paper aims at discussing the mutual synergy of AI and cloud technologies and how cloud solutions support and improve AI. Turning to the core elements of the work, we examine the premises of both AI and cloud computing, their mutual relationship. The paper aims to look into the extent to which cloud infrastructure has transformed the future for the development of AI, particularly with reference to the challenges of scalability and the possibilities of accessing high powered computing, not to mention the added bonus of cost efficiency. New techniques like Federated Learning, AutoML, and AI as a service are shared while contemplating ways of enabling AI to be a breakout of the circle of its enthusiasts. The article also discusses emerging forms of cloud computing, such as edge computing, serverless computing, and hybrid cloud, which will be evaluated concerning the prospects for developing AI and ML. Moreover, it discusses data management within the cloud, including big data analytics, data lakes, and ETL processes optimized for AI. It is important to stress that this article is an attempt to provide a synoptic view of these intertwined subjects in order to explain the change brought by cloud-driven AI and the impacts it can have on technology and society. These findings provide useful information concerning AI and cloud computing to scholars, professionals, and policymakers who are engaging in this growing, dynamic field.

Keywords: *Artificial Intelligence, Cloud Computing, Federated Learning, AutoML, Edge Computing, Data Management*

I. INTRODUCTION

In the quickly changing field of technology, two transformative forces have emerged as key drivers of innovation: artificial intelligence (AI) and cloud computing. The combination of these two promising paradigms has marked the beginning of new generations of computing possibilities in industries and numerous opportunities for solving multifaceted tasks.

With an aptitude to copy some of the human brain's functions and evolve algorithms from a mass of information, AI has become an essential commodity in different realms [1]. This creates the social reality of an ever-growing range of crucial functions being automated and optimized with the help of AI, whether it's diagnostics in healthcare, prediction in finances, etc. At the same time, cloud computing has changed the conditions of storing, processing and utilizing data, providing both horizontal and virtual infrastructures [2].

We have already seen that AI and cloud computing have not happened by chance; it is a perfect partnership where two powerful technologies support each other's strengths [3]. Cloud infrastructure offers counterparts' of computational resources, storage, and networking that AI algorithms need [[4]. Such a partnership helps organizations adopt AI solutions at a larger level and makes the sophisticated tools of analysis and machine learning available in industries on a broader scale. It is imperative to emphasize the role of cloud infrastructure in the development of AI. They afford an environment for fast testing, support group work, and help to meet the variable demand for resources. Furthermore, consumption-based AI services help businesses as well as researchers reduce the entrance fee, contributing to the improvement of the speed of experimentation and development [5]. This article shall therefore seek to look at how the two technologies complement each other and the technologies, methodologies and approaches behind this marriage. This paper shall explore the following subtopics in detail as they relate to the interface between cloud computing and AI: The scalability that cloud computing gives AI. Access to special hardware. Innovative service models. Furthermore, we will explore new trends, including Federated Learning, AutoML, and AIaaS Enabling, with the new technological trend that heralds a new future for AI. In this way, the paper aims to explain the interdependency between these two phenomena and, therefore, shed light on the possibilities opened up by artificial intelligence enhanced by clouds in the future development of technology.

Independent Researcher, Frisco, TX, United States of America

II. LITERATURE REVIEW

A. Understanding the Foundations

Artificial intelligence, or AI, has gone from being a concept seen in movies to becoming part of our reality and significantly changing it. AI can be defined as the areas of computer science focused on creating systems that would be able to solve problems that human intelligence could solve [1]. Some of these tasks are as follows; perception, speech, decision-making, and language translation. The present-day AI has seen a lot of progress in machine learning, but more in deep learning, which has produced breakthroughs in fields such as natural language processing, computer vision and self-driven systems [6].

Today, AI covers a vast number of methods and practices that have been developed concurrently. Supervised learning is a process in which systems can learn from the data labeled for this purpose, while unsupervised learning enables new patterns to be derived from the data not labeled [7].

	Supervised Learning	Unsupervised Learning
Data	Labeled data	Unlabeled data
Learning Task	Predicting or classifying based on labeled examples	Discovering patterns, structures, or relationships in the data
Goal	Generalization to predict labels for unseen data	Extraction of hidden patterns, clusters, or relationships
Algorithms	Decision trees, support vector machines, and neural networks	Anomaly detection clustering, dimensionality reduction,
Use Cases	Image recognition, and fraud detection text, classification, and sentiment analysis	Pattern Recognition in DNA Sequences, Data Preprocessing, and Recommendation Systems

Table 1. Major Differences between Supervised Learning and Unsupervised Learning

Reinforcement learning is used in AI to allow the agents to learn through experience and relationships with the environment [8]. It has grown over the years with new forms like transfer learning and few-shot learning as it tries to optimize machine intelligence.

Cloud computing is a model of delivering computing resources such as servers, storage, databases, networks, applications, analytics, and smart processing through the Internet, or ‘‘the cloud," it can act as a means of getting and using large amounts of computational power as needed, with the principal not necessarily directly and proactively controlling the computational resources [9]. The basic service and deployment characteristics of cloud computing are: self-service, based on the network, pooled resources, instantaneous elasticity, security, pay per use, and metered (see Figure 1) [10].

Key Characteristics of Cloud
On-demand self-service
Measured service
Broad network access
Scalability and rapid elasticity
Security
Resiliency and availability
Pay-per-use pricing
Resource pooling

Table 2. Key Characteristics of Cloud [10]

Cloud services are typically categorized into three main models: IaaS, PaaS, and SaaS [11]. All of these models offer different degrees of control, flexibility and management, so organizations can select the best model that they can use.

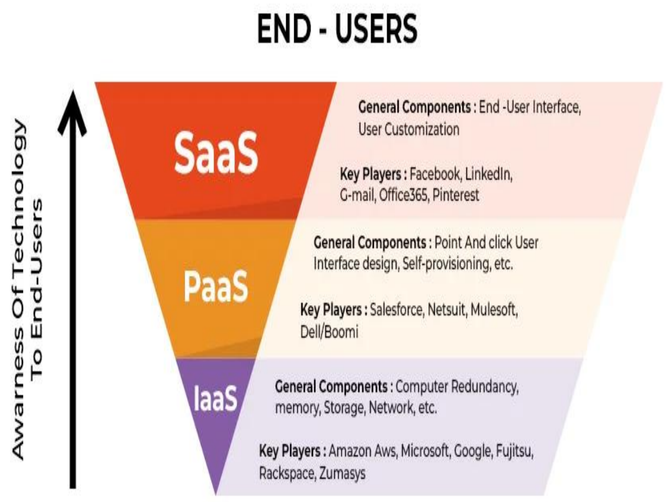

Fig 2. Cloud Computing Services [12]

SaaS is a model through which, instead of purchasing licenses, users can rent software through an Internet browser, with almost any type of software available. PaaS provides computing platforms and is in the middle of SaaS and IaaS in the context of cloud computing. The third type of cloud is IaaS, which provides the actual platform encompassing varied software and hardware on the cloud [12].

It is so because AI and cloud infrastructures are two intertwined technologies, where each thrives off the other’s capabilities. Machine learning and deep learning algorithms that AI encompass call for high computational power and features of data storage and retrieval. These resources are offered in a convenient, elastic, and cheaper way by cloud computing.

This synergy manifests in several ways:

1. Scalability: AI has different varieties of computing tasks; they include training large models or providing results on a cloud platform [13].

2. Accessibility: AI as a Service is therefore a continuous provision of sophisticated AI technologies and AI tools in ways that are not very demanding in terms of capital investment from the consuming organizations.

3. Data Management: Cloud computing, being the fundamental component of its implementation, provides a stable environment for data storage, processing and analytics of the big data required in AI [14].

4. Specialized Hardware: As simple as it sounds, it is imperative to appreciate the fact that cloud providers also supply any resources that may be deemed necessary, such as GPUs and TPUs, which indeed remain one of the most conspicuous service provisions in a hurry to boost the functionality of computations in connection with AI [15].

5. Collaboration: Another significant dependence in AI is the relationship with cloud services because the developers can work together on creating or deploying models at different places.

6. Continuous Improvement: AI models benefit from the fact that cloud systems gather and analyze data in real-time.

Cloud infrastructure in AI creates, implements, scales and brings to market AI applications like never before. This co-dependency has led to improvements in some fields such as healthcare, finance, manufacturing, and entertainment to create smarter devices and systems, hence more adaptable and developed, personalized devices and systems.

B. The Impact of Cloud Infrastructure on AI Development

Cloud infrastructures' influence in the advancement of AI is unprecedented, as they have radically transformed the approach to AI system construction. Laying down at the center of this transformation is cloud, which can be best described as the superior scalability factor and flexibility characteristic of cloud platforms for performing AI algorithms. Another advantage of cloud computing solutions compared to on-site implementation is that it enable the proprietors of AI projects to flexibly control the provision of computing resources in accordance with the demands of a given project [16]. This elasticity is especially important for AI applications because the required CPU resources depend on the stage of the workload's implementation. Through cloud computing, AI techniques have been made available regardless of the identity and size of the organizations to give the best performance needed in the results. Cloud providers provide hardware accelerators for machine learning, like GPUs and TPUs, that are required to train deep neural networks and process big data, respectively [17]. It also allows researchers or developers to try out the new models and techniques of AI without spending large amounts of money on specific hardware.

The cloud has also helped spur the growth of big data storage and processing, which has in turn boosted the development of AI [18]. As for contemporary AI, or better, machine learning and deep learning algorithms, they depend heavily on big data. Cloud platforms offer the avenue to host, handle, and analyze these immense volumes of data conveniently. They provide a variety of data storage services, including object storage for structured data, big data services for maintaining large data sets, and data warehouses for structured data for analytical use in AI systems. Another stunning feature that is often associated with cloud-based AI solutions is the aspect of cost. Pay model is a pay-as-you-go model, which fits perfectly with the large, but sporadic, computations necessary in AI applications. It is about matching the requirements of an organization's training procedures to the need for resources so that certain resources can be duplicated or reduced without affecting performance. This sort of flexibility not only cuts down on capital costs but also provides greater variability and, therefore, quicker cycle times for AI system development and iteration.

III. INNOVATIVE METHODOLOGIES AND APPROACHES

Artificial intelligence as a field is relatively young and is known to undergo constant modifications in terms of methodologies and approaches in order to address the issues concerning the creation and implementation of large-scale artificial intelligence systems. Out of these, Federated Learning, AutoML and AIaaS can be considered as disruptive technologies that are revolutionizing the field of artificial intelligence. These methodologies take advantage of cloud infrastructure in order to remove the limitations that were traditionally observed in the development of such AI and to make the use of complex AI possible for as many entities as possible.

A. Federated Learning

Federated learning has emerged as a unique type of machine learning that solves the problems of privacy and data isolation [19]. Essentially, federated learning is a distributed approach to making AI models learn on decentralized data. Different from the scenario where all the data is stored in a single place, federated learning enables the model to be learned on multiple edge devices or servers holding local samples without transmitting the latter.

The idea that is used in federated learning is focused on the idea of the model taking the data rather than the data coming to the model [19]. It is especially useful when data security is an issue, for example in some medical or financial applications, or when data cannot be accumulated in one place because of legal or purely volumetric reasons. In federated learning, all the nodes that participate in the training process build their individual models, and only the model deltas are transferred to the base server [20]. The nodes then update these models to enhance the overall global model of the central server, which in turn is also sent back to the nodes.

The use of cloud structures offers the following benefits when conducting federated learning. Cloud platforms provide the required computational power and networking means for proper management of the distributed training process. They can safely collect the updates of the local models, coordinate the distribution of the global models, and also offer a large storage capacity for the federated learning framework [21]. Furthermore, cloud services help expedite the implementation of geographically scattered federated learning systems, guaranteeing low-latency communication between the central server and participant nodes.

As seen in Figure 3, federated learning operates on a very basic principle. Each client's unique model is trained using data from smartphones, automobile sensors, bank branches, hospitals, and other sources (cloud data). The model, not the data, is then sent to a central server, which combines both and distributes the updated combined model to each client for additional rounds of updating.

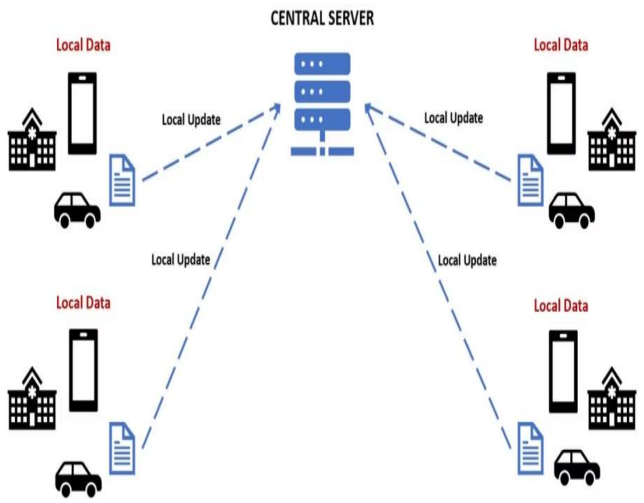

Fig 3. Federated learning—local updates are sent to a central server [22].

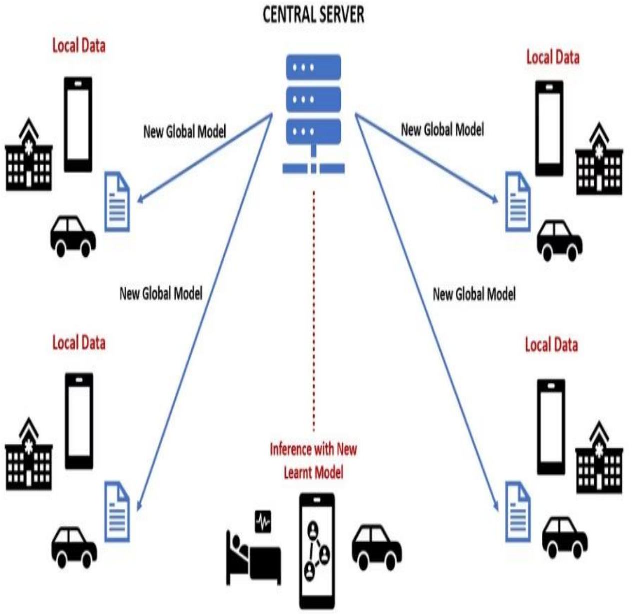

Fig 4. Federated learning: a new global model is dispersed among clients once aggregation takes place on the central server (cloud) [22].

As presented in Figures 3 and 4, the participants (sometimes referred to as clients or nodes) maintain their local datasets with no intention of sharing them with outside parties, while the central server maintains a global model with initial settings.

a. Every participant/client receives the first model from the central server.

b. By training the model that the central server supplied, the participants—who store local data locally or in the cloud—extract knowledge from their local data.

c. The central server receives the trained model back for aggregation (Figure 3), meaning the server simply calculates the average of all the model parameters.

d. The participants receive the updated model from the central server (Figure 4).

All of the clients can thus affect the parameters in the model that underpins the new system. Consequently, one can train a model with all the data while avoiding the collection of the data in the first place. If the model is not convergent, this loop is repeated until the model converges [22]. The count of rounds differs because communication efficiency, processing power, and data distribution are exclusive to each client.

B. Automated Machine Learning (AutoML)

Another concept is automated machine learning, which is an advanced concept that focuses on the automation of the entire ML process to execute on real-life issues. AutoML applies a combination of technologies that relieves a data scientist from performing different tasks of the entire machine learning pipeline, including data preprocessing and feature selection, as well as model selection, hyperparameter optimization, and even model deployment [23]. AutoML can be divided into four types of processes, and in each of them there is a pre-processing stage for automating data acquisition [24]. Preprocessing is also performed by the system: for example, it fills in missing values, encodes categorical features, and scales numerical features. Next, it builds feature engineering, which carries out feature creation or modification to enhance the model's accuracy [24]. At the heart of AutoML are some of the spaces that perform a search through different machine learning algorithms as well as architectures to find the one that proves to be the most efficient for the dataset in question and the problem that it presents.

Another part of AutoML is the process of over-determining the number of hyperparameters [25]. This entails the process of adjusting the coefficients of machine learning models without interference. Hyperparameters are tackled using grid search, random search, and other complex approaches as the Bayesian search methods efficiently explore hyperparameter space. Ensembling is also a common strategy in AutoML systems since it increases the predictive accuracy of models.

Typically, in AutoML, we examine an ML algorithm along with its hyper-parameters. An algorithm is designated as A, and Λ_i represents the domain of its i-th hyper-parameter. Thus, the cross-product of all the hyperparameters in an algorithm with n hyperparameters gives the total vector of hyper-parameters, which has the following structure [25].

$$\Lambda = \Lambda_1 \times \Lambda_2 \times \ldots \times \Lambda_n$$

Vector of hyper-parameters of an algorithm A

Cloud sounding is important in AutoML. They supply the needed computing power for running one or more experiments at a time, thereby making the model search and optimization faster. Cloud-based AutoML services are usually presented as web services with users' friendly interfaces and APIs that let the users upload the data and state the problem, while the service will put together the rest of the AutoML process [26].

Big cloud players are not behind in introducing their own auto-machine learning technologies. For example, AutoML created by Google Cloud contains particular services for the following categories: vision, natural language, and structured data. Amazon Web Services offers automation, and Microsoft Azure also utilizes its AutoML feature in Azure Machine Learning. These platforms make use of the cloud's enormous computational capacity and scalability to facilitate effective AutoML procedures.

C. *AI-as-a-Service (AIaaS)*

AIaaS, also known as AI services, is another approach that is considered revolutionary in terms of how it makes AI available and deployable. AIaaS is defined as the process of availing AI solutions in the form of services via the Internet that enable businesses to harness the power of various AI techniques without having to make capital investments in building up competencies in that area [27]. It should be noted that the primary components of the AIaaS model usually include various AI capabilities provided as cloud solutions. Some of these are pre-trained machine learning models, natural language processing, computer vision, speech recognition systems, and many others. Individuals can use these services via APIs or via a web front-end, which means that the AI functionalities can be readily incorporated into a user's application or process.

The nature and stages by which organizations employ AIaaS are broadly described in the following sections. First, the user has to choose the relevant AI service depending on the need to be met. This could either be a large general; service AI or a niche-service AI with specifically tailored products for specific sectors or specific uses. Then the user passes his data to the cloud platform, ready for analysis, using secure APIs or data-transfer services. The AIaaS platform then analyzes the stated data using the chosen AI models or algorithms and presents the results back to the user. As for the outcomes, it is noteworthy that in many cases the service also offers the tools for performance tracking, mdl selbserv, and expanding the AI-based solution, if necessary.

Cloud providers are also intermediaries performing a critical function in the AIaaS environment. Not only do they supply the backbones for such functions, but they also create and support numerous AI services. These particular services are often constructed with the most advanced AI models and algorithms and are frequently revised by the platform's research departments. This enables users to have the best of advanced AI technology without having to code them or constantly update them on their own.

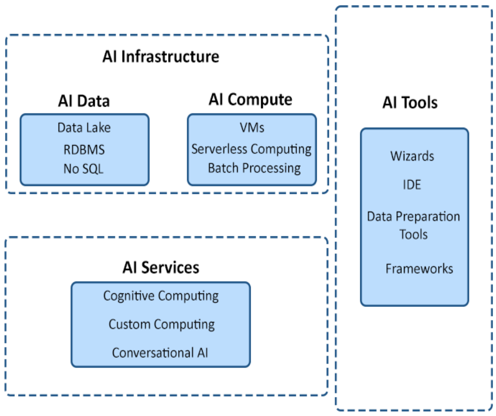

Fig 5. Important Architecture Elements of AIaaS [28]

From Figure 5, AI infrastructure is the way through which AI and ML models are given data and computed as two key components. The derived data is about different databases and data lakes that feed the machine learning procedures to detect patterns and make predictions. New ACC ML algorithms utilize both CPU and GPU for multiplication, which are procured by cloud service providers through AIaaS. With Original Compute Services, AI compute services boost parallelism for VMs and Serverless computing for automating cognitive tasks [28].

Such services as APIs and cognitive services such as speech and texts are also provided through REST endpoints for cloud services. Custom computing enables the training of cognitive services with users' information. One of the components of conversational AI is the inclusion of bots into applications.

Some of the different AI tools offered by these vendors are training wizards, Integrated development environments for managing Machine Learning apps, Data preparation tools for ETL processes, and VMs that are preconfigured with frameworks such as TensorFlow [29]. These tools ease the creation and deployment of smart applications, hence enhancing the application of AIaaS.

There is an assorted selection of AIaaS being provided by major cloud service providers. For example, AWS uses Amazon Rekognition, which is image and video analysis, Amazon Lex, which is a conversational AI interface; and Amazon Comprehend, which is for natural language processing. Similar services, such as vision AI, speech-to-text AI, and natural language AI, are provided by Google Cloud. The cognitive services offered by Microsoft Azure includes speech, language, vision, and decision-making.

IV. NOVEL CLOUD TECHNOLOGIES EMPOWERING AI

The field of cloud computing is always changing, bringing new innovations that are especially useful for applications involving artificial intelligence. Among these, serverless computing, edge computing, and hybrid cloud solutions are notable examples of important developments that are changing the implementation and management of AI systems.

A. Edge Computing

Edge computing has evolved as a strong model for providing computation and data storage closer to the source of data. Closely related to the previous approach, this concept is especially prominent within the context of AI in an application, as it may result in noticeable gains in terms of speed and efficiency when data is processed directly at or near its source [30]. Essentially, the concept of edge computing entails the decentralization of computing assets, which are deployed wherever they would be most effective, which is the periphery or edge of a network.

Based on this point of view, it is possible to mention that edge and cloud computing are complementary solutions for AI applications. Cloud computing brings out the benefits of huge computational power and centralized storage; on the other hand, edge computing is capable of providing real-time processing and low bandwidth utilization [31]. Such integration enables the distributed design of AI solutions in which basic data pre-processing and decision-making can take place on the edge while more resource-intensive activities like analytical computations, model training, etc. can be done on the cloud, if needed.

B. Severless Computing

The serverless system is most appropriate for improving the effectiveness and growth of the AI. It makes it possible for the AI functions to be divided into mini functions that can run in decentralized manners. This fine grained process lets organizations better manage their resources and prevent waste, as resources are utilized only when necessary implementations of serverless AI are numerous. They are; real time image and video processing, whereby serverless functions may be invoked to process an incoming media file. Processing natural language can be done as it serverless, functions, mostly for tasks such as sentiment analysis or language translation, since they can be easily scaled based on the amount of input data processed [34]. Also, there is a nice fit for serverless computing in managing sophisticated AI pipelines, as they imply many models and stages of data processing.

C. Hybrid Cloud Solutions

A hybrid cloud is a mixture of public cloud services and either a private cloud or on-premise infrastructure, so that data and applications can be served between these two [35]. The approach becomes helpful in the case of AI applications in general since a number of them work with sensitive data or have certain performance characteristics. The very concept of a hybrid cloud for hosting AI entails the right blending of an organization's on-premise infrastructure and cloud services. Perhaps the on-premises infrastructure is employed for storing a company's valuable data and training AI models that produce outcomes critical to the business, while the public cloud resources may be utilized for scaling AI applications millions of times or for running AI workloads on custom-designed hardware. In this way, organizations are able to minimize the risks associated with certain parts of the AI processes while at the same time taking advantage of the advantages of using the public cloud, where flexibility and scalability matter. There are several ways in which different approaches to achieving hybrid AI and improving an AI's effectiveness can be managed. One of the approaches is to perform data acquisition and preparation for feature extraction on-premise while lowering the amount of data that is transmitted to the cloud [36]. Thus, it can be used for model training and for carrying out large scale inference. These are:

using the cloud to experiment and for rolling out applications; and using a local or private cloud for the production models due to the need to monitor them closely.

V. DISCUSSION

Several benefits can be derived from applying federated learning with the help of cloud infrastructure. The requirements of the mentioned distributed training setup can be fulfilled by the computational resources and networking facilities of cloud platforms. It can perform the computation of secured model updates, coordinate the replication of the global model, and also offer central storage for the FL system [37]. Also, cloud services help manage obtaining data from geographically dispersed locations while maintaining efficient, low-latency communication between the system's federated learning nodes and the main server [37].

The advantages of such an approach to federated learning are the following: This allows the organization to use the different datasets while at the same time keeping them secure and being able to uphold the protection standards as required by the law [37]. This also means creating models that are less sensitive to the particularities of small amounts of data, collected in one way and no other, which means potentially more general models. Besides, federated learning eliminates the need for the exchange of raw data, thus cutting down on bandwidth consumption and latencies.

But there are also disadvantages to federated learning. It is very important to address the security and privacy issues with model updates that are part of communication between two organizations. There is also a problem of heterogeneity in data across nodes since data is not independent and identically distributed (non-IID), which complicates the convergence of the model [38]. In addition, the federated learning process could be quite time-consuming and complex for the involved devices because it involves optimization of the learning algorithms and communication procedures. Besides, the effect of AutoML on AI democratization and efficiency cannot be overstated. The use cases make it easier to achieve interactivity, which increases the number of people who can create high-quality machine learning algorithms but don't necessarily need to be data scientists. When the capabilities of AI are democratized, they can be used in different fields and by more people [39]. Moreover, AutoML helps decrease the time and efforts needed to train the models and hence accelerate the delivery of machine learning solutions, enabling data scientists to concentrate on other aspects of AI.

When it comes to data management, which is a key part of most AI systems, cloud computing is a game changer for most organizations [40]. Cloud big data and AI integration have provided fresh opportunities for deriving more value and making smarter decisions based on formerly untapped sources of information. Cloud computing enables big data to be analyzed and forms the basis of most artificial intelligence applications [41]. Cloud infrastructure provides services for the storage and computation of big data at scale, speed and in a diverse format. Such capacities facilitate the ability of organizations to analyze datasets that can hardly be processed by traditional on- premise structures. Due to the fact that big data analytical platforms like Apache Hadoop, Apache Spark and others are flexible and easily scalable, they can provide sufficient computing capabilities for AI systems to identify patterns, trends, and correlations in big data.

Application of AI with big data improves the capability of analyzing data sources with relevant information. Machine learning provides the capability to uncover signatures and outliers in large data sets; deep learning handles large unstructured data sets such as images, videos and text [42]. It realizes predictive business analytics, real-time decision-making, and the modeling of learning machines built with big data and AI. Such data storage systems as data lakes and data warehouses become robust solutions for handling the varieties of data needed for AI solutions in the cloud. A data lake is a central gathering place for data, where you can store all kinds of data of any volume and variety. Due to its native format, raw data can be stored and easily used for artificial intelligence and machine learning, which might require raw data to work with.

AI workloads have several benefits when it comes to cloud-based data lakes. It is highly advantageous because it offers a highly portable and scalable form of storage that can work at different capacities and types of data storage [43]. This is important for the kinds of AI applications that may require the processing of various types of data, like sensors and feeds from social media, among others. Also, many cloud data lakes are compatible with big data processing frameworks and AI tools, which simplifies the creation and implementation of AI models. On the contrary, data warehouses contain structured, purposefully filtered data that has been preprocessed for a particular use. Cloud data warehouses provide fast query and analysis functionalities, which make them suitable to be used in business intelligence applications and some applications of artificial intelligence [44]. They can quickly process questions regarding large databases and offer structured data that is often useful for training or testing the artificial intelligence algorithms.

Data preparation and ETL types of solutions (Extract, Transform, and Load) are both important stages of preparing data for AI usage [45]. The cloud platforms' ETL services are rather effective, which enables them to cope with the tasks of data integration and transformation. These services can also include data preprocessing, where things such as cleansing and normalizing the data and determining the features needed can be accomplished.

Managing large data volumes can be easier when done online since such ETL software comes with pre-developed links to numerous data types. They can also work with the extensiveness of the cloud to manage big data in ways that do not take much time. Ironically, even the cloud service providers have prospective solutions that use AI to identify the kind of data, suggest transformation procedures that can be applied, and create ETL workflows based on the input data traits. Some of the preprocessing steps that may be performed in preparing data for use with AI include: dealing with missing values, feature encoding for nominal type data, and normalization or scaling for quantitative type data [46]. Most of these tasks are actually solved by cloud platforms through tools and services that can eliminate these steps and make them less time-consuming for data preparation for AI models. Also, cloud-based tools for data preparation frequently connect with well-known machine learning frameworks, making the transition from data preparation to model training and deployment more efficient.

VI. CONCLUSION

There are several interesting trends and innovations that are already predicted for the further development of AI and cloud computing. Quantum computing finds its application as a cutting edge technology for introducing the A cloud capabilities to its novo level, providing tasks and computation capabilities for complicated AI algorithms. At the same time, there is the notion of green AI to counter the development of even more powerful AI by implementing new energy-efficient models and cloud computing resources.

Another frontier is AI for cloud optimization, where the AI itself is used to optimize the cloud's functions, in terms of resources, energy usage, and efficiency. These advancements thereby reveal the enhancement of cloud infrastructure in AI, and it can be said that cloud infrastructure has made advanced technologies more easily accessible to a broad market and industries.

Thus, the combination of AI and cloud computing holds that the two technologies have initiated massive change in technology. Cloud infrastructure has revolutionized the AI field by offering scalable resources as well as new approaches such as federated learning and AutoMachine learning. The functionality of the cloud in data handling and management cannot be underemphasized, as it has enabled the management of big data, which is useful for AI advancement.

Reflecting on the future, one can expect further enhancement of this cooperation that will result in new discoveries. With growing quantum computing and ever-changing ideas on sustainability, the AI-cloud environment will produce even more potent, effective, and affordable solutions to AI. This ongoing transformation will continue to democratise the use of capacities informed by AI, encourage the evolution of various sectors, and advance humanity's relationship with technology.

References

[1] N. Malik and A. Solanki, "Simulation of Human Brain," pp. 150–160, Jan. 2021, doi: https://doi.org/10.4018/978-1-7998-4763-2.ch009.

[2] K. E. Kushida, J. Murray, and J. Zysman, "Diffusing the Cloud: Cloud Computing and Implications for Public Policy," *Journal of Industry, Competition and Trade*, vol. 11, no. 3, pp. 209–237, Jun. 2011, doi: https://doi.org/10.1007/s10842-011-0106-5.

[3] Wagner, "Cloud Computing and Artificial Intelligence: Opportunities and Challenges," American Cloud, Aug. 18, 2023. https://americancloud.com/blog/cloud-computing-and-artificial-intelligence-opportunities-and-challenges

[4] Klouddata, "Cloud Infrastructure Types, Requirements and Benefits," *www.klouddata.com*. https://www.klouddata.com/infrastructure-blogs/cloud-infrastructure-types-requirements-and-benefits#:~:text=Cloud%20Infrastructure%20Components&text=It%20includes%3A (accessed Jul. 20, 2024).

[5] N.-A. Perifanis and F. Kitsios, "Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review," *Information*, vol. 14, no. 2, Feb. 2023, Available: https://www.mdpi.com/2078-2489/14/2/85

[6] M. Soori, B. Arezoo, and R. Dastres, "Artificial Intelligence, Machine Learning and Deep Learning in Advanced Robotics, A Review," *Cognitive Robotics*, vol. 3, no. 1, 2023, doi: https://doi.org/10.1016/j.cogr.2023.04.001.

[7] S. Palaoglu, "Supervised vs Unsupervised Learning - Explained," *dida Machine Learning*, Jul. 07, 2023. https://dida.do/blog/supervised-vs-unsupervised-learning#:~:text=Summary (accessed Jul. 20, 2024).

[8] J. Carew, "What is reinforcement learning? A comprehensive overview," *SearchEnterpriseAI*, Feb. 2023. https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning

[9] Microsoft, "What Is Cloud Computing? A Beginner's Guide | Microsoft Azure," *azure.microsoft.com*, 2023. https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing#:~:text=Simply%20put%2C%20cloud%20computing%20is

[10] T. Jones, "8 key characteristics of cloud computing," *SearchCloudComputing*, Nov. 03, 2022. https://www.techtarget.com/searchcloudcomputing/feature/7-key-characteristics-of-cloud-computing

[11] Google Cloud, “PaaS vs IaaS vs SaaS: What’s the difference?,” *Google Cloud*. https://cloud.google.com/learn/paas-vs-iaas-vs-saas
[12] N. Soni, “Different Types Of Cloud Service Models In Cloud Computing,” *S2 Labs By Shrey Sharma*, Jan. 12, 2024. https://s2-labs.com/admin-tutorials/cloud-service-models/
[13] Hyperstack, “The Role of AI in Cloud Computing: A Beginner’s Guide,” *www.hyperstack.cloud*, 2024. https://www.hyperstack.cloud/blog/case-study/the-role-of-ai-in-cloud-computing-a-beginners-guide#:~:text=AI%20is%20used%20in%20a (accessed Jul. 20, 2024).
[14] H. A. AL-Jumaili, R. C. Muniyandi, M. K. Hasan, J. K. S. Paw, and M. J. Singh, “Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations,” Sensors, vol. 23, no. 6, p. 2952, Mar. 2023, doi: https://doi.org/10.3390/s23062952.
[15] Zborowska, “Why Generative AI and Cloud Platforms Are a Perfect Match,” *IDC Europe Blog*, Aug. 10, 2023. https://blog-idceurope.com/why-generative-ai-and-cloud-platforms-are-a-perfect-match/ (accessed Jul. 20, 2024).
[16] Salesforce, “Top 12 Benefits of Cloud Computing,” *Salesforce*. https://www.salesforce.com/platform/cloud-computing/benefits/
[17] Silvano et al., “A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms,” arXiv.org, Jun. 27, 2023. https://arxiv.org/abs/2306.15552
[18] M. Brenner, “The Role of AI in Cloud Computing,” *The Forecast By Nutanix*, Jul. 12, 2023. https://www.nutanix.com/theforecastbynutanix/technology/ai-in-the-cloud#:~:text=Investment%20in%20the%20cloud%20is (accessed Jul. 20, 2024).
[19] K. Hu, S. Gong, Q. Zhang, C. Seng, M. Xia, and S. Jiang, “An overview of implementing security and privacy in federated learning,” *Artificial intelligence review*, vol. 57, no. 8, Jul. 2024, doi: https://doi.org/10.1007/s10462-024-10846-8.
[20] Krishnan, “Federated Learning - Abhilash Krishnan - Medium,” Medium, Jun. 30, 2024. https://medium.com/@abhilashkrish/federated-learning-0a2a6485a29f (accessed Jul. 20, 2024).
[21] Y. Shastri, “A Step-by-Step Guide to Federated Learning in Computer Vision,” *www.v7labs.com*, Feb. 03, 2023. https://www.v7labs.com/blog/federated-learning-guide
[22] Karabulut, “Future of AI: Federated learning—the what and why for enterprises,” Jun. 30, 2022. https://www.netapp.com/blog/future-of-AI-federated-learning/
[23] Salehin *et al.*, “AutoML: A Systematic Review on Automated Machine Learning with Neural Architecture Search,” *Journal of Information and Intelligence*, Oct. 2023, doi: https://doi.org/10.1016/j.jiixd.2023.10.002.
[24] M. Baratchi *et al.*, “Automated machine learning: past, present and future,” *Artificial intelligence review*, vol. 57, no. 5, Apr. 2024, doi: https://doi.org/10.1007/s10462-024-10726-1.
[25] S. Herbst, “Automated Machine Learning and Hyperparameter Optimisation,” *Medium*, Dec. 23, 2021. https://medium.com/@sabrinaherbst/automated-machine-learning-and-hyperparameter-optimisation-f33be5eb6901 (accessed Jul. 20, 2024).
[26] M. Barroso, D. Hinjos, P. A. Martin, M. Gonzalez-Mallo, V. Gimenez-Abalos, and S. Alvarez-Napagao, “Boosting AutoML and XAI in Manufacturing: AI Model Generation Framework,” pp. 333–350, Sep. 2023, doi: https://doi.org/10.1007/978-3-031-46452-2_19.
[27] Yasar, “What is Artificial Intelligence as a Service (AIaaS)? - Definition from WhatIs.com,” *SearchEnterpriseAI*, 2023. https://www.techtarget.com/searchenterpriseai/definition/Artificial-Intelligence-as-a-Service-AIaaS
[28] V. Kanade, “What Is Artificial Intelligence (AI) as a Service? Definition, Architecture, and Trends,” *Spiceworks*, 2021. https://www.spiceworks.com/tech/cloud/articles/artificial-intelligence-as-a-service/
[29] TensorFlow, “Data preprocessing for ML: options and recommendations | TFX,” *TensorFlow*. https://www.tensorflow.org/tfx/guide/tft_bestpractices
[30] S. J. Bigelow, “What Is Edge Computing? Everything You Need to Know,” *Techtarget*, Dec. 2021. https://www.techtarget.com/searchdatacenter/definition/edge-computing
[31] Lee, “The Relationship Between Edge Computing and Cloud Computi...,” *www.suse.com*, 2024. https://www.suse.com/c/the-relationship-between-edge-computing-and-cloud-computing/#:~:text=Cloud%20computing%20processes%20data%20in
[32] H. Ali, “Serverless Machine Learning: Building and Deploying AI Models in the Cloud,” Aug. 2023, doi: https://doi.org/10.31219/osf.io/s2u6y.
[33] Dinocloud and M. Vega, “Benefits of Serverless Computing,” *DinoCloud*, Apr. 10, 2024. https://dinocloud.co/benefits-of-serverless-computing/
[34] P. A. Olujimi and A. Ade-Ibijola, “NLP techniques for automating responses to customer queries: a systematic review,” *Discover Artificial Intelligence*, vol. 3, no. 1, May 2023, doi: https://doi.org/10.1007/s44163-023-00065-5.

[35] Raza, “Public Cloud vs Private Cloud vs Hybrid Cloud: What’s The Difference?,” *BMC Blogs*, Aug. 31, 2020. https://www.bmc.com/blogs/public-private-hybrid-cloud/

[36] S. Wang *et al.*, “A Cloud-Guided Feature Extraction Approach for Image Retrieval in Mobile Edge Computing,” *IEEE transactions on mobile computing*, vol. 20, no. 2, pp. 292–305, Feb. 2021, doi: https://doi.org/10.1109/tmc.2019.2944371.

[37] Siniosoglou *et al.*, “Federated Learning Models in Decentralized Critical Infrastructure,” *www.ncbi.nlm.nih.gov*, Jan. 01, 2024. https://www.ncbi.nlm.nih.gov/books/NBK602372/?report=reader#:~:text=FL%20offers%20several%20advantages%20in (accessed Jul. 20, 2024).

[38] Yankai Lv, H. Ding, H. Wu, Y. Zhao, and L. Zhang, “FedRDS: Federated Learning on Non-IID Data via Regularization and Data Sharing,” *Applied sciences*, vol. 13, no. 23, pp. 12962–12962, Dec. 2023, doi: https://doi.org/10.3390/app132312962.

[39] Lawton, “Democratization of AI creates benefits and challenges | TechTarget,” *Enterprise AI*, Jan. 16, 2024. https://www.techtarget.com/searchenterpriseai/feature/Democratization-of-AI-creates-benefits-and-challenges

[40] Adinga, “AI in cloud computing: A game-changer,” *ITWeb*, Aug. 28, 2023. https://www.itweb.co.za/article/ai-in-cloud-computing-a-game-changer/RgeVDqPRwDz7KJN3#:~:text=AI%20in%20cloud%20computing%20represents (accessed Jul. 20, 2024).

[41] “The Innovation Foundation (Part 2): How Big Data Is Related To Cloud, AI and Machine Learning,” *SPK and Associates*, Jun. 23, 2023. https://www.spkaa.com/blog/the-innovation-foundation-part-2-how-big-data-is-related-to-cloud-ai-and-machine-learning#:~:text=Cloud%20computing%20offers%20flexibility%2C%20scalability (accessed Jul. 20, 2024).

[42] M. Taye, “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions,” *Computers*, vol. 12, no. 5, pp. 91–91, Apr. 2023, doi: https://doi.org/10.3390/computers12050091.

[43] Google Cloud, “What is a Data Lake?,” *Google Cloud*. https://cloud.google.com/learn/what-is-a-data-lake

[44] Astera, “What is a Cloud Data Warehouse? A Complete Guide,” *Astera*, Oct. 27, 2023. https://www.astera.com/type/blog/cloud-data-warehouse/#:~:text=Simply%20put%2C%20a%20cloud%20data (accessed Jul. 20, 2024).

[45] IBM, “What is ETL (Extract, Transform, Load)? | IBM,” *www.ibm.com*, Apr. 24, 2024. https://www.ibm.com/topics/etl#:~:text=ETL%E2%80%94meaning%20extract%2C%20transform%2C

[46] “What Is Data Preprocessing for Machine Learning? | Pure Storage,” *www.purestorage.com*. https://www.purestorage.com/knowledge/what-is-data-preprocessing.html#:~:text=Data%20preprocessing%20involves%20several%20techniques