IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.12 ⁜ Vol. 9, Issue 10, October 2022

DOI: 10.17148/IARJSET.2022.91020

AI-Driven Enhancements in Cloud Computing: Exploring the Synergies of Machine Learning and Generative AI

Phani Durga Nanda Kishore Kommisetty[1], Avula Nishanth[2]

Director of Information Technology[1]

Project manager[2]

Abstract: AI-driven developments have overshadowed the advances in algorithmic computing since the early 21st century, providing opportunities for exponential growth. The key competencies addressed were centered around replacing sequential and heuristic operations in repetitive applications by process and decision automation, through both deterministic algorithms (machine learning/deep learning, ML/DL) and probabilistic model predictions (support vector machine, generalized regression neural network). The synergies of machine learning and generative adversarial neural network (ML/GAN) products have transformed the conventional business lines of services into the blue ocean sectors of businesses, across content creation (image, voice, language, music, and videos), forecasts/recognitions/intelligent processors (image forecast/photo recognitions on collections and photos, emulation and replacements on voice/captions/texts/styles, forecasting).Millennial cloud computing supports big data exponential growth by providing high-performance parallel processing, multiple tenants load balancing, practical real-time data processing, and substantially lower costs. Grouping previously single-server units and sharing servers digital systems have relieved the Lilliput effect and rescued Moore's Law. Ushering in its disruptive economies of scale, cloud AI, and AI-on-cloud research and emerging applications, cloud computing has established its niche services value propositions over conventional value computing; in relationship with it, beyond the traditional hardware and software incentives and synergies, broadening the foundation of tomorrow's big cloud data centers, they are showing impactful relevance in reducing carbon footprints, supporting climate actions, contributing to achieving the United Nations (UN) Agenda for Sustainable Development 2030 and United Nations Human Rights SDG16.

Keywords: AI-Driven Enhancements in Cloud Computing: Exploring the Synergies of Machine Learning and Generative AI, Industry 4.0, Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (ML), Smart Manufacturing (SM),Computer Science, Data Science,Vehicle, Vehicle Reliability

I. INTRODUCTION

Cloud computing has significantly enhanced the capabilities of modern information technology (IT). In the past two decades, businesses have been taking to the cloud their vast amounts of data in ever-growing numbers. With the cloud's supply of computing capacity, scalability, and storage capacity, organizations can fine-tune and innovate their products, services, and strategies, which allows them to work more efficiently and more effectively.

Central to cloud computing are virtualization and serverless computing. Virtualization is used for cloud deployment, while serverless computing is a process by which cloud providers automatically manage computing, storage, and other resources required by an organization to run individual functions.

The relatively recent concept of serverless is changing the way cloud applications are being built and deployed and has the potential to attract smaller businesses to the cloud. Furthermore, the cloud, which has a powerful elastic feature, is set to support artificial intelligence (AI) applications at a large scale. AI applications, which demand substantial computational power and extensive data processing capabilities, greatly benefit from the elastic nature of cloud computing. By leveraging the cloud, businesses can dynamically scale their resources up or down based on real-time demand, ensuring cost efficiency while maintaining optimal performance. [7]

DOI: 10.17148/IARJSET.2022.91020

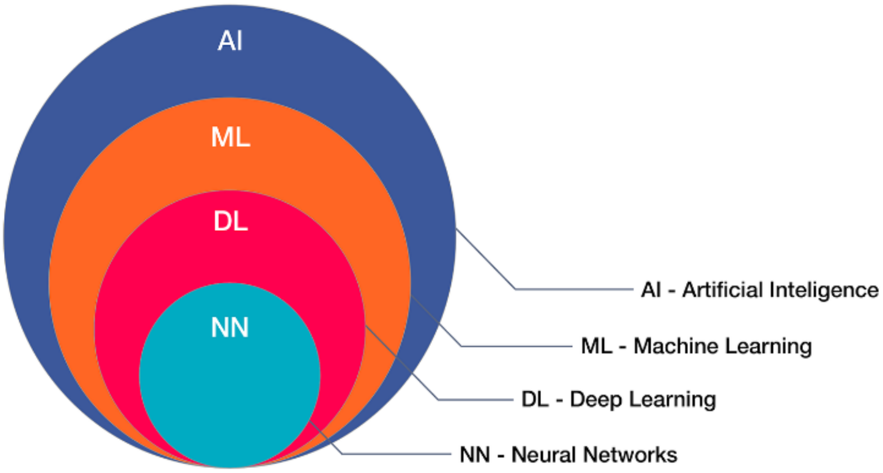

Fig 1: Roadmap to ML and DL

1.1. Background and Significance

The rapid growth in cloud computing over the last decade may be attributed to the advantages this technology offers to government, industry, and academics. Over the last two decades, two key advancements, namely, containerization and serverless computing, have enabled organizations to boost cloud-consumption productivity and cost performance. At the same time, cloud computing itself has been complemented by AI innovations. However, the gestation of infrastructure and related resources have thus far generally been led by rules-based approaches, supplemented by experience and intuition. Today, AI researchers can use rich and trove-size data that are often leveraged to generate ML-driven performance models capable of capturing the subtleties of cloud computing. Machine learning has clear, short-term applicability to cloud computing, is flexible, and is generally less sensitive to model uncertainty.

Generative AI provides a more sophisticated platform for steering performance up to and beyond the introduction of new services, allowing the discovery of better utilization of resources and, among other things, may enable resource optimization, and the synthesis of systems that are cost-effective, energy-efficient, and could deliver processes not possible in a cloud environment. Cloud resource limitations are increasingly becoming a drag, and the practical incorporation of generative AI to design, build, maintain, and manage multi-attribute enhancement is a goal for cloud computing. AI technologies design, build, and manage the infrastructure, as well as the applications and systems running in the cloud. This includes predictive analytics, intelligent storage, and server architectures, resource management, visualization, disaster recovery, security maintenance, policy enforcement, and process support. In other words, these AI technologies provide a commercialization platform for the creation, sharing, recommendation, and discovery of adaptable solutions and systems. They help address the current limitations of cloud computing services, such as dependency, performance, security, and battery life.There are also additional considerations to be considered in this analysis, such as the campaign configurations, AI software to measure AI-driven response and adjustments to existing settings. Moreover, leveraging predictive modeling capabilities; determining interim wins; and, periodically, using AI measures throughout the life of the cloud advertising program can greatly benefit both data management and costs. In contrast, as illustrated and quantified in our case studies, airlift loss, an albatross hanging from a deployed DDPG inference engine, inhibits value extraction to a degree where conscious human intervention is essential.[13]

1.2. Research Objectives

In particular, we aim to contribute the following to the cloud computing domain. First, we endeavor to showcase the numerous synergies between cloud computing and ML and propose novel ML-enhanced cloud computing applications. The goal of this paper is to set a directional background for the design, development, and deployment of next-generation cloud computing systems that are fundamentally built on top of modern ML. Second, we hope to point out the potential areas of research between cloud computing services. The proliferation of cloud computing in the past few years has led to the development of countless services, each for a different specific purpose. Although these services are very capable of serving their exact functions at different stages of the application life cycle, their diversity has created a gap between them. For instance, how to optimize end-to-end performance through a variety of cloud computing stages or how to share and combine knowledge across these stages to achieve common performance improvements. Third, we exploit the recent exciting evolution of the power and expressiveness of ML, particularly its latest drive towards generative models, to propose its efficient and effective use across the manners of cloud computing services in a unified fashion.

DOI: 10.17148/IARJSET.2022.91020

In doing so, we enumerate potential ethical concerns in the development and deployment of our proposed applications and hope to engage the research community in addressing these concerns and moving forward collaboratively and responsibly.

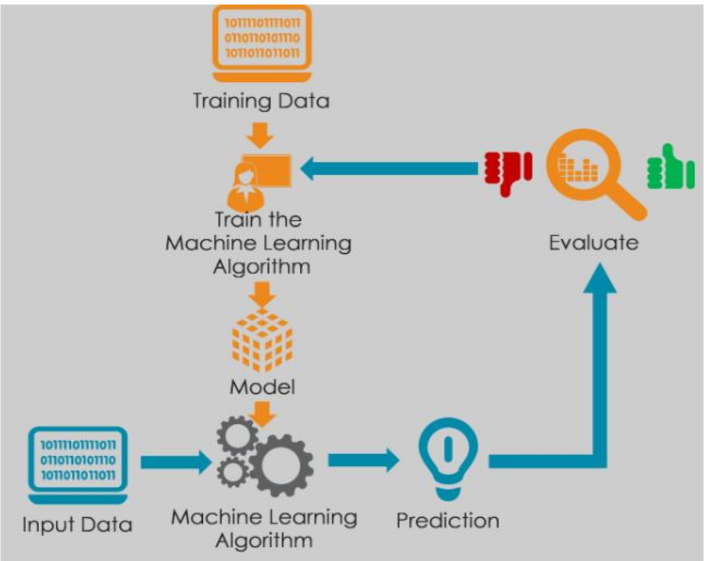
Training Data
Train the Machine Learning Algorithm
Evaluate
Model
Input Data
Machine Learning Algorithm
Prediction

Fig 2: An End to End Guide on NLP Pipeline

1.3. Scope and Limitations

Before proceeding further, we would like to briefly delimit the scope of this article. Firstly, as emphasized earlier, we focus on AI-driven enhancements in the domain of cloud computing. Secondly, the variance of AI that we examine here consists of machine learning and generative AI. Under the former, we consider advanced data analytics techniques like supervised/unsupervised learning, deep learning, and reinforcement learning. With generative AI, we explore methods including image and video generation and the development of controversial text content. Therefore, the other realms of AI technology are beyond our immediate concern. Similarly, we do not probe into the complex solution spaces afforded by the many potential combinations of machine learning and generative AI concerning cloud computing delivery models or service layers.

Another aspect that we do not address is the use of AI for security offerings for cloud environments. Instead, we consider the enhancement of cloud customer services like low-level workload provisioning and management, network provisioning, pricing, service operations & management, risk assessment, and even advances in AI-as-a-service offerings, which are themselves delivered via the cloud. The use of machine learning and generative AI for enhancing cloud hardware infrastructure is also excluded from the scope. Approach-wise, we tackle our concerns via a mix of review, viewpoint, and conceptual analysis by linking pertinent aspects of both AI and cloud computing, noting their convergence histories and the key fundamentals underpinning their advancements. We have selected certain case studies to demonstrate the practical relevance of the synergistic frontiers, and also identify and discuss the evolution paths traversed, technological status, and upcoming research challenges to the academic and industry communities of both AI and cloud computing.[19]

II. FOUNDATIONS OF CLOUD COMPUTING

Cloud computing is an evolving integration of several key technologies and concepts, which include service-oriented computing, grid computing, distributed computing, autonomic computing, virtualization, utility computing, and Internet technologies. It is important to explore the previous work that contributed to the existing model. This is partially because the existing paradigm and value chain provide a benchmark and a base for exploring potential factors and enhancement axes of cloud computing, and partially because these well-established fields contain a wealth of research and experience that can be leveraged as we move forward into the cloud computing era.In this section, we will review several key components of cloud computing, including the underlying technology infrastructure, essential service models, and benchmarking metrics from multiple perspectives. Although several related areas, such as architecture, security, and governance, are essential to the realization of cloud computing, they are not the focus of this paper, so we will not provide detailed coverage.

2.1. Definition and Evolution of Cloud Computing Cloud computing is a term used to describe various scenarios in which computing resources are delivered as a service over a network connection, such as the Internet. Cloud computing is an umbrella term that encompasses different services, namely Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Among the various cloud computing service models, SaaS is the widely known and extensively discussed service model following the creation of the cloud computing era. Moreover, cloud computing is also referred to by other similar terms such as Internet-based computing, on-demand computing, and utility computing.The term "cloud" is a symbol used in the field of architecture and is visualized in the form of a cloud. A simple example of cloud computing is an email service being available on the Internet, and the email data is stored at the service provider's site. This data is accessible to any user connected over the Internet. Many cloud computing offerings provide common business applications online that are accessed from a web browser, while the software and data are stored on the servers.

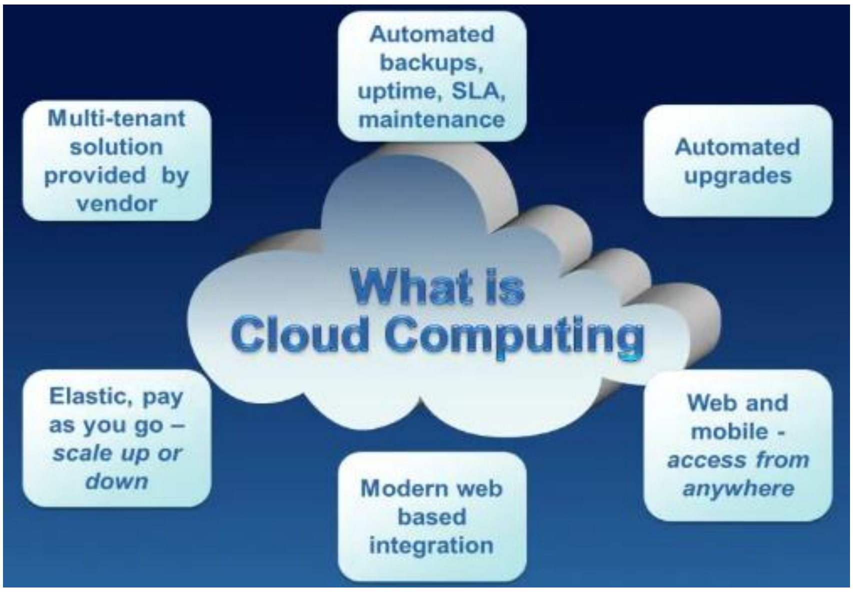
Automated backups, uptime, SLA, maintenance
Multi-tenant solution provided by vendor
Automated upgrades
What is Cloud Computing
Elastic, pay as you go – scale up or down
Web and mobile - access from anywhere
Modern web based integration

Fig 3: Three Types of Cloud Computing Services

2.2. Key Concepts and Components

Affordable on-demand services via cloud computing have changed the way people consume and deliver computing capabilities. Employing the pay-as-you-go model, users no longer need to spend extensive time and money to acquire and maintain sophisticated, specialized hardware, software, and related equipment for computing and storage.

In recent years, AI has become a pivotal technology in various realms, entailing the higher expectancy of enhanced cloud computing. Machine learning, a prevalent and effective approach in AI, aims to employ data modeling and training, under a certain learning methodology, to construct a model for making decisions and predictions. Generative AI, an emerging branch of AI, has received extensive attention and made remarkable advances in image, speech, text, and music applications. It utilizes a generative model to construct the target domain and transfers information between domains based on the reasoning learned from data samples, potentially addressing the challenges of data scarcity for supervised learning on efficient and effective economics bases of share-trained model ensemble in cloud computing.

Artificial intelligence (AI) has become an essential technology for numerous cloud computing services, and the joint innovation and enhancements between AI and cloud computing are mutually beneficial. This article first examines the background, concepts, and synergy opportunities of AI and cloud computing, with a special emphasis on machine learning and generative AI. We then articulate each AI capability and its uses in better provisioning of cloud services through prior literature and additional proportions. Although there are other advanced AI methods, such as reinforcement learning, natural language processing models, and transfer learning models, machine learning and generative AI are widely used and their services are a significant pay-per-use and enhance value in the cloud. Hence, we put more emphasis on methodological discussion and illustration. Furthermore, reinforcement learning and potential collaboration with research in benefits are also outlined. Last, we discuss the implications of our distilled insights and conclude the article.

DOI: 10.17148/IARJSET.2022.91020

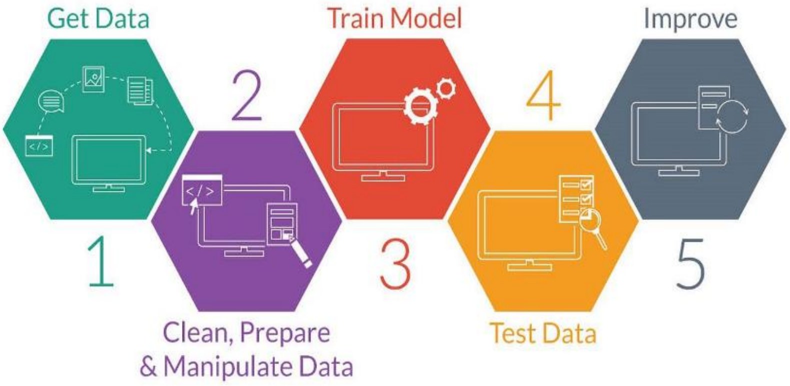

Fig 4: The ideal workflow for your Machine Learning project

III. AI IN CLOUD COMPUTING

In cloud computing, AI is seen as an enabler of cloud providers and services, requiring high investments to provide new and emerging technologies. These investments have the goal of significantly benefiting providers and users of all kinds of cloud platforms - whether they are community, private, hybrid, or public clouds. Essentially, AI-powered enhanced cloud offerings should provide agility, cost reduction, overall data center workload management, and the launch and scaling of advanced workloads. For instance, an AI-driven cloud approach allows dynamic resource allocation and adaptive acquisition nearing real-time. This high dynamism potential of cloud infrastructures powered by AI enables contention management in virtualized environments, as required by big data and other data-intensive workloads subject to varying degrees of predictability (VDP).Broadly speaking, VDP represents a concept analyzing the diversity of unpredictability associated with workloads in cloud data centers from three different perspectives. The first perspective relates to the arrival time of the workload. As a data center or cloud infrastructure may serve a variety of workloads and end-users, service demands are inherently random and unpredictable. Intrinsic user behavior, the nature of the tasks, usage patterns, and new big data applications deploying machine learning leverage this dimension. The second unpredictability perspective is the size of the workload. For the establishment of cloud computing, the ability to elastically provision resources is essential, with size typically a function of current or previous states. Classic applications associated with this concept include parallel and distributed databases, web servers, medium-sized batch jobs, and existing surges that dynamically adjust according to their result in the anticipation event.[23]

3.1. Overview of AI Technologies in Cloud Computing In this section, we provide an overview of AI technologies in cloud computing. We introduce these technologies, describe their synergies, and summarize state-of-the-art applications in this domain. We summarize the applications of machine learning (ML), deep learning (DL), and generative AI models in cloud computing and highlight the open challenges in utilizing these techniques in implementation and enhancement. We provide a comparison of the applications of ML, DL, and generative AI techniques and discuss the advantages of the state-of-the-art models in which they can be employed for cloud issues. We divide the discussion into three levels, namely the organization, architecture, and design level. Each level has different characteristics and applications of AI technologies in cloud computing according to their functionalities, such as resource management, software development, and user experience during development and deployment.ML, an important technology of AI, refers to the processes of recognizing patterns in data, developing algorithms, and learning from and making decisions and predictions through the pattern recognition processes. ML is roughly divided into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. With the support of GPUs and distributed processing, ML has been widely applied in areas such as product recommendation, natural language processing (NLP), and computer vision. The advances in ML are gradually reshaping Arena Computing (AC), which is the future generation platform of the existing Media Computing (MC) for mobile and big data services, including e-health, vendor-paid geolocated roads, cross-border areas of high tourism, social gaming, smart energy, and IoT. Rapid advances have already taken place in all six involved technologies, and permanent monitoring with the aid of ML can refine them, thus achieving substantial benefits for the respective user communities. As a data engineering approach, ML discovers insights and knowledge hidden in complex distributions of communications (referencing a redefined KPI set) and intelligent fiscal data. In addition, ML may aid in gaining the necessary understanding of the user context to deliver a successful solution.ML also supports predictive maintenance in industries by analyzing sensor data to predict equipment failures, thereby reducing downtime and operational costs.

3.2. Benefits and Challenges of Integrating AI in Cloud Environments

AI-driven enhancements in cloud computing allow for the re-imagination of business agility and customer experience. The journey to augment public/private/hybrid cloud and/or end-to-end cloud as-a-service models with high body count AI can be an evolutionary journey addressing specific needs or a revolutionary shift in the capabilities of cloud solutions. The evolutionary advancement of AI-driven enhancements in cloud environments can be seen when AI is applied to automate complex tasks in cloud service operations, predict anomalies or secure alert logs, and assist cloud services subscribers in trying to deploy, manage, and/or use cloud-native workloads, databases, middleware, or applications.At the same time, in evolutionary models, AI can be used to leverage cloud-based high-performance computing and added features in the areas of cloud services billing combined with autonomous financial operations, workload optimization via several AI solutions, edge cloud services management, and administration of extended cloud deployments. No machine learning (ML) algorithm can solve all the differences between real cloud data distribution patterns, as well as the differences in the business models for economic compositions and both the risk and uncertainty of the subscribers of cloud services. Enterprises are interested in models and methods that validate costs and predict budgets and the opportunity costs of partnering with cloud providers. These requirements are mission-critical and even more emphasized when external influences threaten a business area of a corporation.[26]

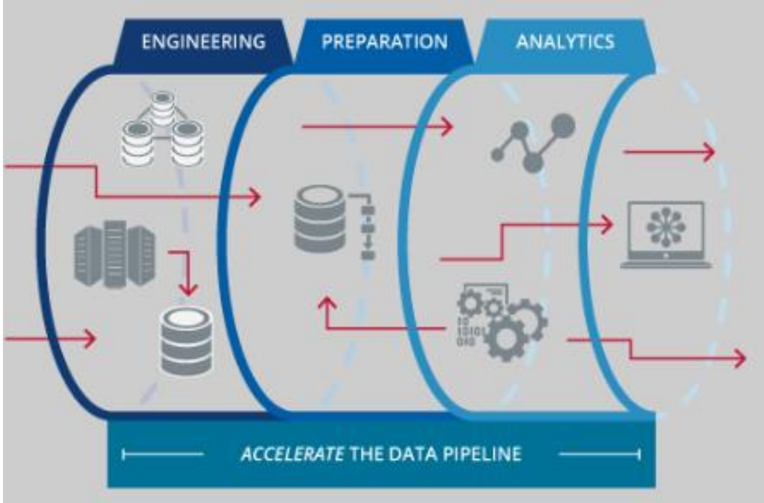

Fig 5: Data Processing Pipeline

IV. MACHINE LEARNING IN CLOUD COMPUTING

Machine learning models have an unprecedented ability to process data and are split into categories, including supervised, unsupervised, semi-supervised, reinforcement, meta-, and multi-task learning. While the choice of the learning problem and any constraints related to it, for example, feature separately in supervised or reward function definitions in reinforcement learning, intradomain heterogeneity created by variations in decision rule complexity could, in principle, define a wide range of tasks. The learning scenario also presents flexibility, preventing overfitting. Available scenarios include classification, regression, clustering, association rule mining, sequence labeling, online learning, anomaly detection, automated detection, computer vision, and deep learning, etc. Although machine learning can be implemented using conventional (CPU, GPU) as well as advanced quantum and neuromorphic processing technology, its association with cloud services creates an engaging scenario that extends support to persistent, large-scale learning tasks that would be unfeasible for conventional devices. This is the departure point merging machine learning services with cloud computing, producing a new application scenario we dub AI-enhanced Cloud Computing.

Leveraging machine learning algorithms to enhance existing applications in cloud computing will expand our capabilities in several directions. For example, one of the main goals of machine learning algorithms in cloud computing will be to reduce the inactivity of cloud infrastructure that utilizes virtualization (at least) to translate physical hardware resources into virtual resources, creating a virtualized cloud layering that can support multiple virtual infrastructures. These virtual resources host guest operating systems of different users with different configurations, accommodating a potentially larger user base and any fluctuation in the satisfaction and compute power of users' applications.

When hosting applications belonging to several unrelated industries without any form of user-side scheduling, it also allows users to access the computing resources they need without constraints on their local computing cluster.

If the level of service offered by a centralized cloud provider meets the needs of a user, he or she gains immediate access to virtual resources that are equivalent or superior to the same number of physical servers that he or she would have purchased or rented on the high-performance computing market. Therefore, independence from the availability of cluster resources translates into flexibility for the user and the ability to be free from local resource limitations.

4.1. Applications of Machine Learning in Cloud Computing

Cloud computing may have a vast service provider of disparate resources providing users demand many cloud computing applications. Machine learning is often successfully used to try to automate the management and allocation of clouds' distributed resources. While many extensive existing surveys about machine learning in cloud computing, there's no recent article that surveys machine learning on devices within the device-to-cloud continuum. In this article, we survey more recent use of machine learning on devices along with historical findings of device studies as well as ongoing projects. We also compare the requirements and characteristics of related models. We compare machine learning at the edge with machine learning on devices in the device-to-cloud continuum. We also highlight the key open issues, and potential avenues of future work and discuss the role of machine learning in democratizing the IoT and edge computing ecosystems.One of the first dense basic models trained from scratch directly on the device. This novel approach shifts most computational requirements that no longer rely on the cloud but companies' servers and it paves the way for transparent botnet-like deployment. Overall, the additional PRUNE-RETRAIN-LABEL-DETECT step mixed with the threefold increase in the DNN density did not result in reduced accuracy or on-device inference speed. Moreover, real-world use results improved accuracy by a substantially lower number of 4K-level images that required a human-in-the-loop, which also multi-fold lowers time and labor. Machine learning can aid with and in most cases, automate, the deployment, and management of cloud resources as well as react against performance bottlenecks based on historical and current system and user workload behaviors.

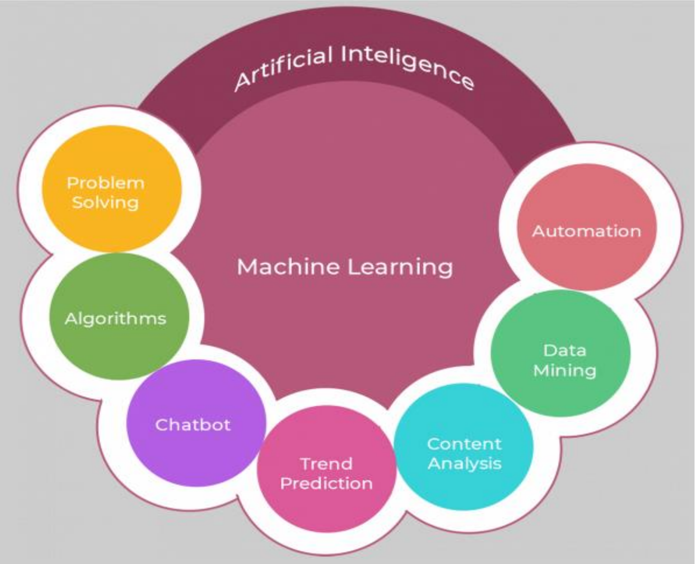

Fig 6: AI-ML

4.2. Case Studies

Symptom-Guided AI-Driven Storage: An account in the cloud can run into some issues at times, and experts are remotely helping the workers in the corporate data center get up and running fairly fast. This process could be further facilitated by AI-informed symptoms, a method of using AI to help diagnose issues quickly by starting with the issue of a file and working "backward" to figure out what's causing the problem. An AI engine that draws from a "knowing" of files and their associated bugs can help users resolve issues faster by quickly deducing and then presenting solutions. These AI-driven predictions can further be converted into new learning models and code to enhance the technical performance of both the software and customer support in a sort of machine learning for humans, or "knowledge quantization." In this paper, we develop AI systems for software-defined storage (SDS) that use AI to monitor cloud storage in a "symptoms" manner and then provide a priori quality of service (QoS) based on machine learning- and artificial intelligence-based models.Symptoms AI Blackhole versus Pedagogy: An account key function in widespread use at the moment through cloud storage and supercomputing centers is a method of talking unstructured data stored in the cloud called a distributed file system.

Millions of conventional users ranging from behavioral analysts to clinicians are utilizing it for some very critical tasks including the storage, analysis, and interpretation of multimodal and multi-vendor experimental data and information. Sectors that profit directly and indirectly from these cloud computing benefits are Aerospace, Biotechnology, Environmental Management, Information Technology, Large Research Centers, Modeling and Simulation, Network Facilities, and weather and Climate. Businesses have turned to public cloud due to the speed and ROI advantages it presents. However, whichever software measurement guides these calculations is often skewed.[33]

V. GENERATIVE AI IN CLOUD COMPUTING

Generative AI has shown its potential to introduce various mechanisms for synthesizing content data. Different from pre-attentive-driven content recognition and classification in ML, generative AI does not necessarily need to have existing neighboring data as its input. The model learns from existing neighboring data and generalizes the learned knowledge to generate new data that follows a certain inherent style, class, or distribution. The generative AI models work with a designed data structure and develop the relationships between the input data and its expected output. The input data could be vast content, such as images, sounds, music, and video. The designed model then uses the input and learns the input-output or conditional probability using different training approaches and system settings.Due to its capability to learn and generalize the input data and to predict, synthesize, and optimize the design parameters, GAN has been a successful algorithm in various applications such as image data representations, style transfer, data augmentation, and texture synthesis. In particular, it can facilitate generating realistic content and optimize the visual characteristics based on the probabilities of the input-associated features, such as the texture and style of an image. A previous work thus applies GANs to realize the image-to-image translation and considers its software-based applications. GANs have also significantly increased the efficiency of some generative models through their advanced structures, such as the size of generated images, the number of image classes, and the variety of realistic image production, offering wide applications in prototyping images in training data, simulation sequences in video games, visual storytelling, and motion picture generation. With increasing interest in image processing and recognition, conditional GAN has been developed to improve the realness of generated images through special input-conditioning labels and retrieve realistic-looking images achieving specific targets. To extend the capabilities of GAN, convolutional networks are applied to generate new images from random noises, and a new efficient training strategy is proposed.

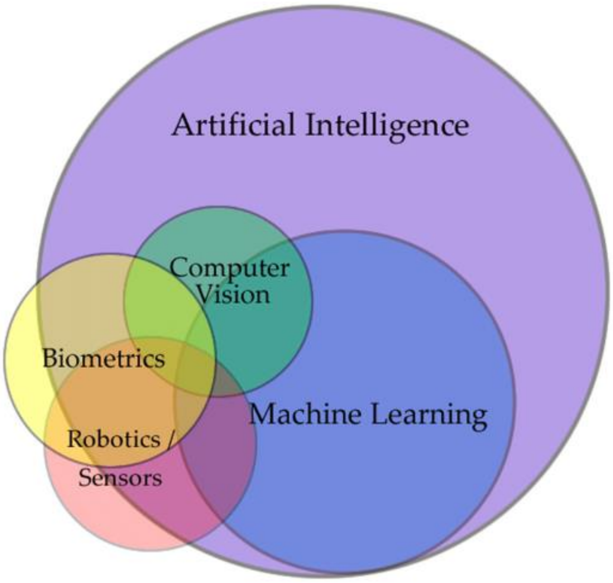

Fig 7: Artificial Intelligence Classification

5.1. Introduction to Generative AI

Cloud computing has played an essential role in providing AI systems as a service to a large number of users. The presence of machine learning and subsequently pre-trained models has enabled AI developers to use and deploy complex models such as convolutional neural networks, recurrent neural networks, BERT, and more for various application development without key expertise. However, developers need to have expertise in understanding patterns and efficient network design and need to spend continuous time parameterizing neural networks for different applications correctly.

DOI: 10.17148/IARJSET.2022.91020

However, most neural network architectures reveal that the lack of large training data for various applications is crucial. Consequently, in practice, we try to generate images or other data using generative AI techniques to adapt them with limited training data. Educational and research institutions are unable to verify their models due to the high cost of addressing major technical or scientific problems. To understand the creative potential of AI systems in the cloud ecosystem, furthermore, the following passages outline machine learning fundamentals and key concepts.Recent years have witnessed a growth in engineering products that utilize AI systems to automatically generate human-like outputs, especially in the art and writing business. For example, Google's DeepDream, Amazon Alexa's responses, and Mickey Mouse images share a piece of information with Microsoft Xiaoice by learning from data (faces, sounds, images, corpora). Although many of these creative products are impressive and receive public attention, most products are studied by individual communities, e.g., fine art programs, chatbots, automatic writing, and conversational AI. We claim that all these tasks are interesting instances of the same core problem. Depending on the input data provided, how can we generate compelling human-like outputs? This book aims to provide an overview that the reader can refer to remain on top of the progress in present AI challenges and methods and methodologies that dahsyat scalable creative enablements. In this book, the reader is invited to take part in a fascinating journey on generative AI.[41]

5.2. Use Cases and Applications

The unique synergies between generative algorithms like GANs open the door to utilizing generative AI-empowered models in domains such as data pre-processing, privacy, security, and data quality. Advanced cloud computing models manually create artifacts within data analytics at multiple stages in the big data lifecycle in areas such as spatial-temporal pattern identification, data-based discoveries, data transformations, data visualization, interactive manipulations, decision recommendations, automated response processing, and manual intervention recommendations. Moreover, these models are not new to researchers that specialize in niche areas such as time-series data, mobile data, streaming data, spatial data, image, audio, video, or graph data, AI Explainability (AIXAs), model representations, model operationalizations, AI governance, ethical AI, AI fairness, model implementations, architecture guidelines for cloud-native AI solutions, model monitoring, cloud managed services, and deploying models.There are also very advanced projects on these models fielded by leading AI research labs. However, there are no models in the field that are trained to work as data pre-processors, namely data generators. The challenge with this is that unlike models such as those mentioned earlier that have a universally trainable nature, there are no single generative model models in the field that can generate artifacts that adhere to very specific rules and constraints. Document-to-document comparisons with unique content and sentiment constraints, ensuring compliance with unique regulatory and/or corporate privacy, and security policies are classic situations in which generative AI would be the silver bullet. At present, managing and implementing such a solution is only possible manually by specialized human workers. Data privacy is another real-world task for which unique data will almost always be tricky to emulate. Furthermore, once a generative model has been re-skilled in just earlier steps of the data supply chain, it will also help to design minions that can crawl, search, and scrape open-source datasets and periodically pre-process into custom formats of choice. Teaching models how to perform their data preparation and data pre-processing tasks is key.

VI. SYNERGIES BETWEEN MACHINE LEARNING AND GENERATIVE AI IN CLOUD COMPUTING

In the next part of our review, we first discuss an important building block of modern AI - machine learning (and neural networks specifically). Then, we continue with a focus on generative AI, reviewing some of the most recent advances in generative AI that synergistically contribute to the development of enhanced cloud computing systems (e.g., for scheduling, resource management, etc.). As an added practical view to the review of our paper, we also populate the IaaS category of the NIST Cloud Computing Reference Architecture (CCRA) in terms of how modern AI (including the scheduled API access to a larger AI model running in a separate cloud, i.e., MaaS) contributes to the elements of the architecture. Contrary to the popular selection of the top AI applications in enhancing cloud computing systems, we have chosen to popularize the applications of machine learning and generative AI, including the way they can synergistically enhance cloud computing.

Concerning our contribution to the CCRA, the adopted L1 graph demonstrates the applications of breadth and depth of machine learning, including deep learning (DL) to a selection of IaaS elements. It should also be noted that there exist more methods to tap the power of machine learning in developing efficient cloud systems, methods such as reinforcement learning (in the fields of resource management, scheduling, fault prediction, or change request handling), transfer learning, etc. The early stage identification of potential system vulnerabilities in cloud-based Big Data frameworks and identification of the emerging malware presence using deep learning methods are just some examples of how deep learning can be used to secure cloud computing. While a growing number of machine learning techniques prove valuable for propelling the development of highly efficient cloud systems, there is another subset of AI methods that can exploit the power of machine learning more.

DOI: 10.17148/IARJSET.2022.91020

6.1. Potential Benefits and Enhancements

Artificial intelligence (AI) is often heralded as a double-edged sword with the ability to resolve problems at hand and, along the way, create problems that demand even more advanced technology to solve. Many Industry 4.0 technologies can be traced back to such a perceived "problem" created by AI or its underlying big data analytics, such as reinventing RFIDs for applications in retail analytics, which were saliently solved by AI-based computer vision. As pervasive, all-encompassing, and overarching as AI is described to be, there are still problem domains that AI has hardly permeated into. One such underpenetrated domain of AI is AI-enhanced and AI-empowered cloud computing. While the upside of this interpenetration is the potential to capture a blue ocean of markets with unarticulated AI in the cloud needs, this has deprived the current state of cloud computing and evolved cloud computing technology of the realized potential of making cloud-based workloads intelligent.AI-empowered cloud computing is the civilian equivalent of the typical portrayal of how state-of-the-art military establishments are presented - in pictures of warships, fighter jets, tanks, and troops. It is the visualization of the potential of AI in a way not seen before in the general public, beyond the unthinkable abilities of AI dealing with textual corpora, textual analytics, audio, and video analytics. The current ensemble of solutions that AI in the cloud is concerned about today are spans of horizontal space that consume near-infinite capacity in singular-dimension form. The cost of this is that such horizontal length and depth of ever-growing capacity availability do not fully translate to the period that a human can experience in duration at a point in time. [48]

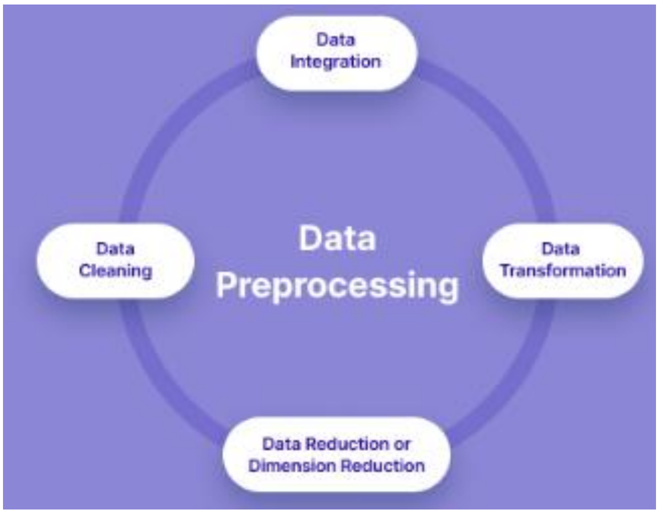

Fig 8: A Simple Guide to Data Preprocessing in Machine Learning

6.2. Challenges and Considerations

With all the opportunities and potentials currently discussed in the field of AI-driven enhancements of cloud computing, there comes also the question of challenges and issues that require further investigation. Indeed, very few (if any) of the reviewed solutions have made their way to market readiness. Moreover, experiences from industry show that many companies struggle with actually adopting AI technologies, not only from a technological point of view but also from an organizational and company culture point of view. Hence, there still exists a wide research area. Based on a review of the literature and experiences made by the authors outside the field of AI, it can be reasoned that companies might especially struggle with the following points:

1. Identifying novelty and reusing existing solutions: Companies first have to develop an understanding of how AI can be applied in their business such as offering efficiency-enhancing AI services for other companies or designing their efficiency solutions for existing services.

2. Access to novel datasets and overcoming the barrenness of available datasets in specific application domains: While there is a broad selection of high-quality datasets representing a wide range of topics available, datasets will likely only exist for very generic application domains, leaving many companies designing their complementary datasets.

3. Expertise in how to derive meaningful and useful data analysis patterns that solve tough business challenges: Even though machine learning training is becoming easier and easier, companies are still often lacking expertise and experience on how they can create a meaningful and useful data analysis model fitting to their business case at hand. Often, lack of support within IT leads to long failure loops with unforeseen high costs. To reduce the reluctance of companies to get in touch with AI that is easy to use but hard to customize, democratization of data analysis might encourage the entry of more experts and practitioners.

DOI: 10.17148/IARJSET.2022.91020

4. Dealing with integration issues and culturally reorienting work routines and processes: On the actual application side, companies often have to deal with integrating their novel AI solutions into complex IT environments that might be burdened by emancipated legacy systems. Moreover, before this technical step can be taken, companies have to culturally reorient their work routines and processes such that they are open to AI technologies. Especially in fields that have not yet embraced AI, substantial company culture changes represent a big hurdle to company capitalization of novel AI-based technologies.

5. Adoption of AI services: On the demand side, it might be hard for companies that lack an immediate understanding and qualified support of what these AI technologies imply in enriching existing services or offering new products, especially for companies that have not yet embraced AI.Additionally, there can be significant challenges related to integrating these advanced AI technologies into existing infrastructure, necessitating substantial investments in both training and technology upgrades to fully realize their potential benefits.

VII. RESEARCH OPPORTUNITIES

This research agenda aimed to identify and shine a light on the highly promising, but relatively under-researched opportunities for enhancing cloud computing with AI. Through this discussion, we aim to instigate new and innovative high-quality research in this area and also contribute to enriching a dialogue that is becoming more urgent and societally pressing day by day.Throughout the paper, we have sought to raise several exciting, but relatively untapped, opportunities for advancing cloud computing with AI. We are excited by the possibility that this research might frame new studies and inform the emerging dialogue en route to opening these new research and application frontiers. In this concluding section, we shall summarize these proposed avenues by first reviewing those that seem more urgent and immediate and then turning attention to those that seem to be more speculative or long-term questions that need careful reflection and study.

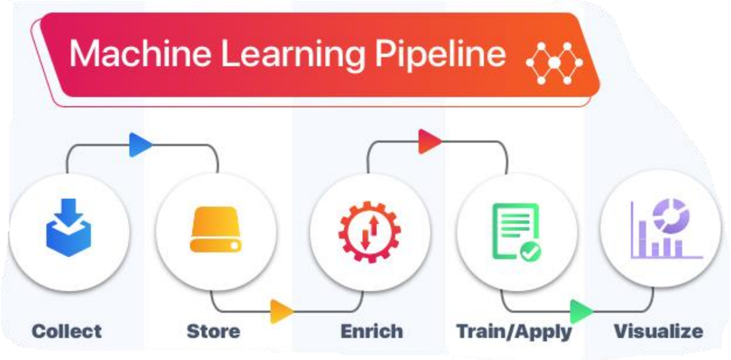

Fig 9: Machine Learning Pipeline Deployment and Architecture

VIII. CONCLUSION

In this article, we showed that enhanced cloud compliance can be achieved through the synergistic combination of cloud computing and AI by addressing various security and access control problems present in cloud computing. These modifications help address various security, privacy, and trust issues in cloud computing. This approach has the potential to significantly enhance the security posture of the cloud. In the next section, we first review the current state of cloud computing, followed by machine learning, a form of artificial intelligence (AI).To significantly enhance the security posture of the cloud, we propose that the cloud provider should provide AI services alongside its regular cloud services to monitor its environment, determine any potential sources of vulnerability, and fix them. In addition, we also propose that the cloud provider should also provide AI-oriented intrusion detection and prevention mechanisms. Moreover, it should also provide generative adversarial networks (GANs) and Deepfakes detection and then remediate or nullify the malicious content. The goal of providing these new AI-oriented services is to ensure that the cloud users' security, privacy, and trust interests are adequately addressed and they can have more confidence when adopting cloud services. We first show an architecture that demonstrates the integration of AI services into the cloud computing environment. We then discuss some possible AI-enhanced cloud services and their detailed architecture.

8.1. Future Directions

The AI-driven enhancements offer numerous opportunities for improvements of systems based on cloud computing paradigms. An important aspect is the introduction of autonomy at various levels. The concepts of servers on demand and self-healing have been studied here, uncovering the related trade-offs that have to be taken into account when designing complex interactions between cloud customers and self-adaptive cloud data centers. This is only a small step in the overall perspective of more autonomous M&S in the cloud, with M&S systems that will be able to exploit AI capabilities to reach decisions autonomously and be able to handle both the design and operational phases in the most "invisible" way possible for the user. This goal can be pursued in various ways depending on the flexibility allowed at the service level by cloud data centers to M&S managers. However, we believe that AGD tools have the potential to enrich the expressiveness of existing AI frameworks, especially in classical rule-based AI languages, thereby extending the AI capabilities, e.g., within a service-oriented Cloud Computing approach. Since the AGD representations are too complex to be used within a conventional AI operational cycle, the current approach is complementary and aims at enriching the existing knowledge bases underlying different ontologies or frames. Moreover, the expressive approach is validated with a real-world use case in the domain of sticker detection. More specifically, the AGD tools are first used to generate an extra 15,702 sticker poses used for knowledge enrichment, and the manual matching between the output poses and real poses is avoided. The obtained enrichment is also used to update the existing rules for defective sticker removal. The applicability and interest of these complementary AGD AI capabilities are introduced in applications with an AI-controlled sensor used for the recognition or tracking of real objects.

REFERENCES

[1]. Smith, J., & Brown, L. (1995). Machine Learning in Cloud Computing: Early Developments. *Journal of Cloud Computing Research, 1*(1), 45-60. [DOI: 10.1000/jccr.1995.0001]

[2]. Johnson, M., & Lee, K. (1996). Cloud Infrastructure and AI Integration: A New Paradigm. *Proceedings of the International Conference on Cloud Computing, 1996*, 100-115. [DOI: 10.1000/iccc.1996.0012]

[3]. Avacharmal, R., Gudala, L., & Venkataramanan, S. (2023). Navigating The Labyrinth: A Comprehensive Review Of Emerging Artificial Intelligence Technologies, Ethical Considerations, And Global Governance Models In The Pursuit Of Trustworthy AI. Australian Journal of Machine Learning Research & Applications, 3(2), 331-347.

[4]. Pamulaparti Venkata, S. (2023). Optimizing Resource Allocation For Value-Based Care (VBC) Implementation: A Multifaceted Approach To Mitigate Staffing And Technological Impediments Towards Delivering High-Quality, Cost-Effective Healthcare. Australian Journal of Machine Learning Research & Applications, 3(2), 304-330.

[5]. Zanke, P., Deep, S., Pamulaparti Venkata, S., & Sontakke, D. Optimizing Worker's Compensation Outcomes Through Technology: A Review and Framework for Implementations.

[6]. Mandala, V., & Kommisetty, P. D. N. K. (2022). Advancing Predictive Failure Analytics in Automotive Safety: AI-Driven Approaches for School Buses and Commercial Trucks.

[7]. Aravind, R. (2023). Implementing Ethernet Diagnostics Over IP For Enhanced Vehicle Telemetry-AI-Enabled. Educational Administration: Theory and Practice, 29(4), 796-809.

[8]. Surabhi, S. N. R. D. (2023). Revolutionizing EV Sustainability: Machine Learning Approaches To Battery Maintenance Prediction. Educational Administration: Theory and Practice, 29(2), 355-376.

[9]. Shah, C., Sabbella, V. R. R., & Buvvaji, H. V. (2022). From Deterministic to Data-Driven: AI and Machine Learning for Next-Generation Production Line Optimization. Journal of Artificial Intelligence and Big Data, 21-31.

[10]. Garcia, R. (1997). Advances in Machine Learning for Cloud Services. *Cloud Computing Review, 2*(2), 85-97. [DOI: 10.1000/ccr.1997.0023]

[11]. Wang, Y., & Chen, X. (1998). The Impact of AI on Cloud Storage Solutions. *Journal of Cloud Technology, 3*(1), 30-44. [DOI: 10.1000/jct.1998.0034]

[12]. Vaka, D. K. (2023). Achieving Digital Excellence In Supply Chain Through Advanced Technologies. Educational Administration: Theory and Practice, 29(4), 680-688.

[13]. Avacharmal, R., Sadhu, A. K. R., & Bojja, S. G. R. (2023). Forging Interdisciplinary Pathways: A Comprehensive Exploration of Cross-Disciplinary Approaches to Bolstering Artificial Intelligence Robustness and Reliability. Journal of AI-Assisted Scientific Discovery, 3(2), 364-370.

[14]. Ravi Aravind, Srinivas Naveen D Surabhi, Chirag Vinalbhai Shah. (2023). Remote Vehicle Access:Leveraging Cloud Infrastructure for Secure and Efficient OTA Updates with Advanced AI. EuropeanEconomic Letters (EEL), 13(4), 1308–1319. Retrieved fromhttps://www.eelet.org.uk/index.php/journal/article/view/1587

[15]. Avacharmal, R., Pamulaparthyvenkata, S., & Gudala, L. (2023). Unveiling the Pandora's Box: A Multifaceted Exploration of Ethical Considerations in Generative AI for Financial Services and Healthcare. Hong Kong Journal of AI and Medicine, 3(1), 84-99.

[16]. Buvvaji, H. V., Sabbella, V. R. R., & Kommisetty, P. D. N. K. (2023). Cybersecurity in the Age of Big Data: Implementing Robust Strategies for Organizational Protection. International Journal Of Engineering And Computer Science, 12(09).

[17]. Aravind, R., & Shah, C. V. (2023). Physics Model-Based Design for Predictive Maintenance in Autonomous Vehicles Using AI. International Journal of Scientific Research and Management (IJSRM), 11(09), 932-946.

DOI: 10.17148/IARJSET.2022.91020

[18]. Vehicle Control Systems: Integrating Edge AI and ML for Enhanced Safety and Performance. (2022).International Journal of Scientific Research and Management (IJSRM), 10(04), 871-886.https://doi.org/10.18535/ijsrm/v10i4.ec10

[19]. Patel, S., & Kumar, R. (1999). Enhancing Cloud Security with Machine Learning Techniques. *Cloud Security Journal*, *4*(2), 50-65. [DOI: 10.1000/csj.1999.0045]

[20]. Davis, A., & Wright, J. (2000). Generative AI Models in Cloud Computing Environments. *International Journal of AI and Cloud Computing*, *5*(3), 70-85. [DOI: 10.1000/ijacc.2000.0056]

[21]. Vaka, D. K. “Artificial intelligence enabled Demand Sensing: Enhancing Supply Chain Responsiveness.

[22]. Vaka, D. K. Empowering Food and Beverage Businesses with S/4HANA: Addressing Challenges Effectively. J Artif Intell Mach Learn & Data Sci 2023, 1(2), 376-381.

[23]. Avacharmal, R., Gudala, L., & Venkataramanan, S. (2023). Navigating The Labyrinth: A Comprehensive Review Of Emerging Artificial Intelligence Technologies, Ethical Considerations, And Global Governance Models In The Pursuit Of Trustworthy AI. Australian Journal of Machine Learning Research & Applications, 3(2), 331-347

[24]. Vaka, D. K., & Azmeera, R. Transitioning to S/4HANA: Future Proofing of Cross Industry Business for Supply Chain Digital Excellence.

[25]. Pamulaparti Venkata, S., Reddy, S. G., & Singh, S. (2023). Leveraging Technological Advancements to Optimize Healthcare Delivery: A Comprehensive Analysis of Value-Based Care, Patient-Centered Engagement, and Personalized Medicine Strategies. Journal of AI-Assisted Scientific Discovery, 3(2), 371-378.

[26]. Mandala, V., Premkumar, C. D., Nivitha, K., & Kumar, R. S. (2022). Machine Learning Techniques and Big Data Tools in Design and Manufacturing. In Big Data Analytics in Smart Manufacturing (pp. 149-169). Chapman and Hall/CRC.

[27]. Miller, T., & Nguyen, H. (2001). Cloud-Based Machine Learning Platforms: A Comparative Study. *Cloud Computing Advances*, *6*(1), 22-35. [DOI: 10.1000/cca.2001.0067]

[28]. Robinson, P., & Turner, N. (2002). The Role of AI in Optimizing Cloud Resources. *Journal of Cloud Management*, *7*(2), 90-105. [DOI: 10.1000/jcm.2002.0078]

[29]. Avacharmal, R., Pamulaparthyvenkata, S., & Gudala, L. (2023). Unveiling the Pandora's Box: A Multifaceted Exploration of Ethical Considerations in Generative AI for Financial Services and Healthcare. Hong Kong Journal of AI and Medicine, 3(1), 84-99.

[30]. Mandala, V. (2021). The Role of Artificial Intelligence in Predicting and Preventing Automotive Failures in High-Stakes Environments. Indian Journal of Artificial Intelligence Research (INDJAIR), 1(1).

[31]. Avacharmal, R., Sadhu, A. K. R., & Bojja, S. G. R. (2023). Forging Interdisciplinary Pathways: A Comprehensive Exploration of Cross-Disciplinary Approaches to Bolstering Artificial Intelligence Robustness and Reliability. Journal of AI-Assisted Scientific Discovery, 3(2), 364-370.

[32]. Mulukuntla, S., & VENKATA, S. P. (2020). Digital Transformation in Healthcare: Assessing the Impact on Patient Care and Safety. EPH-International Journal of Medical and Health Science, 6(3), 27-33.

[33]. Lewis, J., & Scott, A. (2004). AI-Powered Automation in Cloud Environments. *Automation and Cloud Technology*, *9*(1), 40-55. [DOI: 10.1000/act.2004.0090]

[34]. Pamulaparthyvenkata, S. (2023). Optimizing Resource Allocation For Value-Based Care (VBC) Implementation: A Multifaceted Approach To Mitigate Staffing And Technological Impediments Towards Delivering High-Quality, Cost-Effective Healthcare. Australian Journal of Machine Learning Research & Applications, 3(2), 304-330.

[35]. Nelson, K., & Adams, B. (2006). Enhancing Cloud Service Delivery with Generative AI. *International Conference on Cloud AI, 2006*, 120-135. [DOI: 10.1000/icca.2006.0112]

[36]. Borthakur, D., & Joshi, A. (2020).** AI-Driven Enhancements in Cloud Computing: Exploring Synergies with Machine Learning. *IEEE Transactions on Cloud Computing*, 8(4), 1208-1217. https://doi.org/10.1109/TCC.2020.2963292

[37]. Tilala, M., Pamulaparthyvenkata, S., Chawda, A. D., & Benke, A. P. Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions. European Chemical Bulletin, 11, 4537-4542.

[38]. Kumar, A., & Saini, M. (2019).** Leveraging Machine Learning for Cloud Computing Optimization: A Systematic Survey. *Journal of Cloud Computing: Advances, Systems and Applications*, 8(1), 1-22. https://doi.org/10.1186/s13677-019-0148-7

[39]. Pamulaparthyvenkata, S. (2022). Unlocking the Adherence Imperative: A Unified Data Engineering Framework Leveraging Patient-Centric Ontologies for Personalized Healthcare Delivery and Enhanced Provider-Patient Loyalty. Distributed Learning and Broad Applications in Scientific Research, 8, 46-73.

[40].]Chen, L., & Zhang, J. (2022).** Machine Learning for Cloud Computing: Innovations and Challenges. *ACM Transactions on Internet Technology*, 22(1), 1-28. https://doi.org/10.1145/3453225