

Lenovo TruScale – The Right Infrastructure for AI as a Service (AlaaS)

Introduction

Across every industry, organizations are implementing artificial intelligence (AI) to accelerate innovation and improve their competitive position, the quality of products/services, operations, and customer engagement. While AI's promise and economic value are immense, so are the challenges of implementing AI, given the large datasets and the need to effectively store, analyze and protect all of your valuable data throughout its lifecycle.

This whitepaper discusses these deployment and infrastructure challenges and the limitations of traditional on-premises high-performance computing (HPC) infrastructure and public clouds. It then outlines how new, more agile service-based models can allow organizations to leverage new opportunities and accelerate their digital transformation to the next level.

The Remarkable Promise of AI

Machine Learning (ML), a subset of AI, can increase value for many business cases. Examples include identifying product defects, speeding response time to earthquakes by analyzing seismic data, reducing inventory shrinkage in retail, developing autonomous systems such as drones, robots, and self-driving cars, increasing customer retention and up-/cross-selling opportunities, and more.

Consequently, AI adoption is increasing. According to a recent [McKinsey survey](#)¹ in 2022, AI adoption in organizations is 50%, double what it was five years before. In addition, "AI high performers" companies report higher revenue growth than their peers who lag in AI deployment. According to the [Goldman Sachs Macro team](#)², AI adoption could create about \$7 trillion in global economic growth over ten years.

However, although many enterprises have started adopting AI, several find the process challenging, and a few find it impossible. Deploying AI and the associated high-performance information technology (IT) infrastructure can be complex and expensive, especially for Deep Learning (DL), a type of ML that uses Neural Networks to simulate human decision-making.

Deployment and IT Infrastructure Challenges for AI

Deploying AI in production, especially DL workflows, typically has four stages (Figure 1)³:

1. **Data Management** to prepare data needed to build the ML model.
2. **Model Learning** to define, select, and train the ML model.
3. **Model Verification** to ensure the model meets specific functional and performance requirements.
4. **Model Deployment** to integrate the trained model into the IT infrastructure and run, maintain, and update the model as needed.

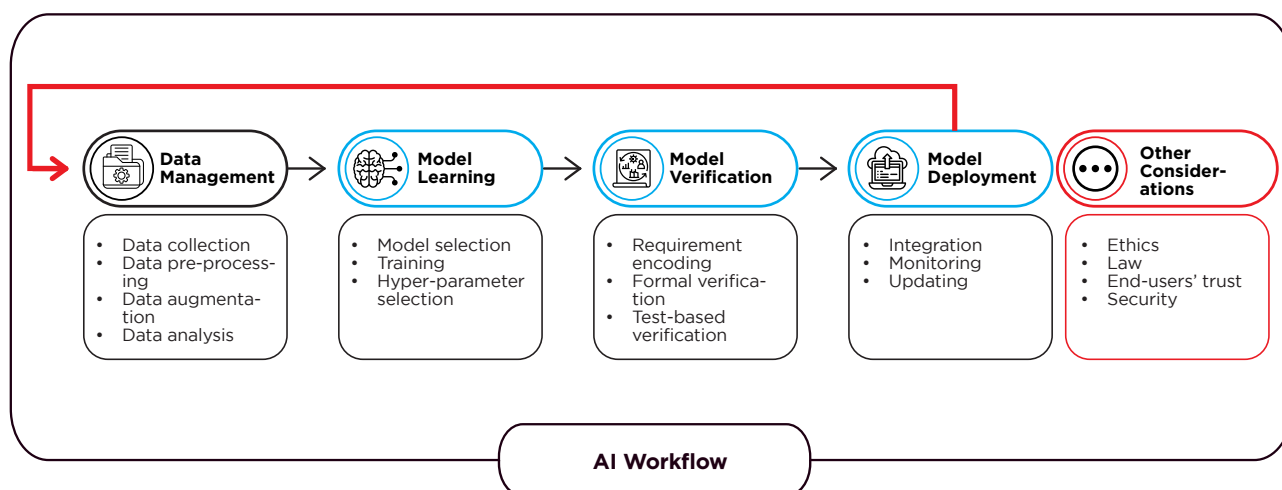


Figure 1: The AI workflow and other considerations

These stages have smaller steps (listed in Figure 1) that can run in parallel and with feedback. In addition, there are other ethical, legal, trust, and security considerations. All of this makes AI deployment very challenging.

While there are many business, skills, and process challenges in deploying AI at scale in production, the IT infrastructure must be well-designed, agile, enterprise-grade, and scalable. This requires high-performance storage and servers optimized for every stage of the AI workflow, along with high-value solutions and services to deploy and operate this IT infrastructure:

High-performance storage: AI workflows iteratively process vast amounts of unstructured data in a short timeframe – an essential attribute since large and often rapidly changing data sets are required to deliver accurate algorithms. This drives significant scalability, performance, flexibility, and affordability storage requirements, often needing high-performance parallel file systems and maintaining data gravity (keeping storage close to the compute to minimize data movement costs). In addition, this storage must have enterprise-grade features that ensure security, compliance, and data resiliency.

High-performance servers: The Model Learning and Model Verification steps are complex, iterative, and time-consuming. For instance, these steps can take days/weeks for a model with a billion parameters unless properly optimized and scaled. To complete these steps in a few hours, one typically needs high-performance clusters (often with accelerators such as graphics processing units or GPUs) and scalable data and model parallel algorithms that distribute the ML/DL computational kernels over thousands of processors.

Organizations must repeat these steps (Figure 1) to experiment with different neural-network topologies, algorithms, and hyper-parameters to reach the desired accuracy level and validate the results before Model Deployment, which may require Edge servers (often with GPUs) to get timely results.

High-value solutions and services: Some key solutions and services include tailored, proven, and ready-to-deploy AI solutions for end-to-end operations, including computer vision, audio recognition, prediction, security, and virtual assistants for every industry. These solutions and services help organizations implement AI more easily by overcoming resource limitations and challenges with high-performance infrastructure deployment and operations.

IT Infrastructure Challenges and Requirements for Deploying AI in Production

There are many challenges (Figure 2) in deploying a high-performance infrastructure for AI:

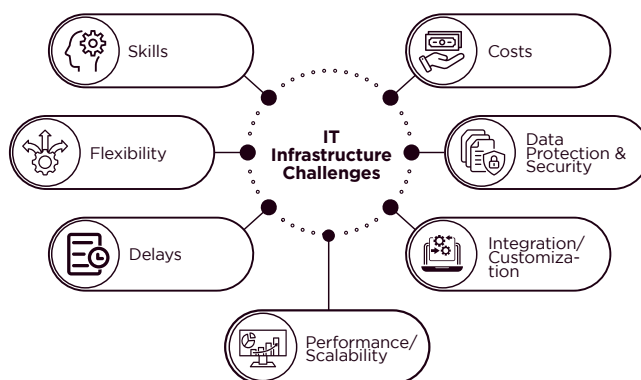


Figure 2: IT infrastructure challenges for deploying AI

- **Costs:** The capital cost of acquiring a traditional on-premises high-performance infrastructure can be prohibitively high, especially for smaller organizations. Additionally, these systems may have lower and sub-optimal utilization under normal operations since they are sized to handle peak loads. These systems also require considerable electricity, backup power, data center space, and cooling, which can add even more costs.
- **Skills:** Operating an on-site high-performance infrastructure is complex and time-consuming. It needs a highly skilled team of technical professionals, which may be challenging to recruit and retain.
- **Performance/Scalability:** DL requires high-performance clusters, hardware accelerators, faster storage, and proximity to data. In addition, each step of the production AI workflow must be optimized for performance.
- **Flexibility:** Applications developed for single high-performance cluster will likely not work optimally on another system. Migrating the entire AI environment onto a new high-performance infrastructure can take weeks or months.
- **Delays:** High-performance infrastructure is in high demand across industries and is complex to install. The procurement, delivery, and deployment time can slow down the pace of implementation.
- **Integration/Customization:** As a rapidly evolving field, customers must customize and integrate newer algorithms and frameworks into their existing AI workflows.

- **Data Protection and Security:** Many organizations deal with sensitive personal information in highly regulated environments and must ensure that this data is protected and secure.

In the future, these infrastructure challenges will increase in complexity as organizations grapple with several thorny challenges while managing, processing, and analyzing higher volumes of data with greater efficiency.

They need flexible, scalable environments that rapidly process tremendous amounts of data to avoid these obstacles and speed up innovation. They must also collaborate and share large data sets and results with upstream and downstream partners. And they need an affordable infrastructure that supports automation to simplify data aggregation, assimilation, and management.

AlaaS provides the infrastructure to overcome many IT and deployment challenges for organizations deploying AI in production.

Overcoming AI Infrastructure Challenges with "as a Service" Models

The growth of a new generation of AI applications in computer vision, natural language processing, recommender systems, autonomous vehicles, drug discovery and genomics, financial modeling, etc., is increasing the scale and demand for high-performance infrastructure. This infrastructure's scale and costs for these AI use cases are so colossal that even the largest organization may be unable to install and operate it without using an "as a Service" or cloud-like model.

With AlaaS, users can acquire and access a range of bare-metal clusters (dedicated single-tenant physical servers with no operating system overheads) from a central pool in a hosted data center (Figure 3 – left) that could be in a cloud or on-premises or be co-located. They can build resource-intensive applications with the option to expand their footprint in the future if required. In addition to raw computing power, AlaaS (Figure 3 – right) typically includes software for cluster management, workload optimization, security, and highly skilled people resources to operate the AI clusters and ensure they are fully optimized to deliver maximum performance for AI applications.

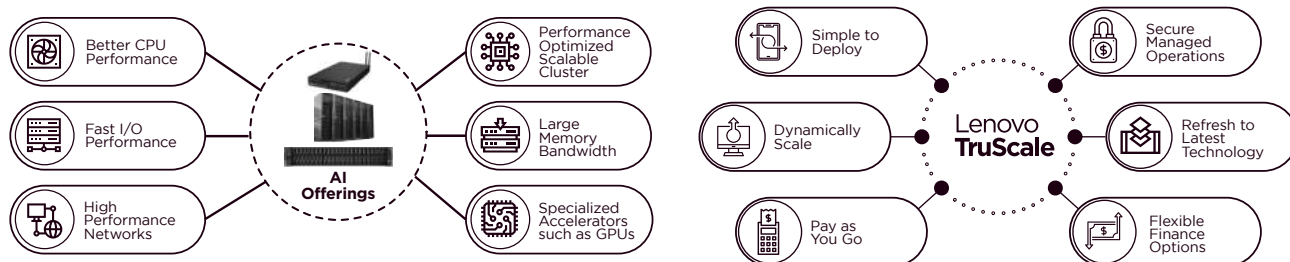


Figure 3: Characteristics of AI Infrastructure offerings and benefits of "as a Service" models

AlaaS helps users save on capital spending and still realize their business objectives. Since the service provider manages all the complexities of the computing environment, customers can get the best value from their investment and focus all their precious resources on improving research and discovery. However, public clouds have many limitations.

Overcoming Public Cloud Limitations

Cloud computing has become an increasingly popular tool for AI for its ability to store, manage, and analyze large amounts of data. However, there are several limitations of cloud computing for AI (which multiply with the number of users and the size of the data files), including:

1. **Operating Cost:** While cloud computing offers scalability and flexibility, it can be expensive, particularly for large-scale AI projects that generate massive amounts of data. In addition, cloud providers charge additional fees for accessing, storing, and using the stored data. These budget constraints can limit access to cloud computing resources for smaller organizations with limited funding.
2. **Data Privacy and Security:** Storing large amounts of sensitive data in the cloud raises concerns about privacy and security. The data must be protected from unauthorized access, theft, or breaches, no matter where it resides. What are the costs associated with protecting the data in the cloud? Is data stored redundantly to ensure recoverability? How much control does the user have over remote storage?
3. **Regulatory Compliance:** Most enterprise IT solutions must be validated for compliance with regulators. The applications and the installed systems environment must be validated on the cloud. With localized IT infrastructure, this is assured by the internal IT group. However, the cloud provider must also do this in the cloud. This typically adds additional costs, risks, and complexity, especially during some compliance audits.
4. **Usability:** An AI workflow involves a pipeline of scripts and a command line interface. Setting up nodes in the cloud requires users to understand the system's intricacies to manually set up these command lines and scripts. In addition, sourcing and optimizing specific applications needed for the entire AI workflow is cumbersome and sometimes not feasible. All this is expensive and time-consuming and may have to be repeated for each user type in the organization.
5. **Internet Connectivity/Data Transfer Times:** Cloud computing requires high-speed internet connectivity to access and analyze large datasets. Internet connectivity can be a limitation for users working in remote or low-resource settings, where internet connectivity may be unreliable or unavailable. In addition, uploading and downloading large datasets to and from the cloud can take significant time and resources, particularly for users with limited bandwidth or slow internet connections.
6. **Technical Expertise/Dependence on Cloud Providers:** Cloud computing requires technical expertise to set up and manage the infrastructure and tools necessary for AI. Lack of technical expertise can be a limitation for users who may not have the skills or resources to use cloud computing resources effectively. In addition, cloud computing relies on third-party service providers, which can limit the users' control over their data and analysis tools. This can raise concerns about vendor lock-in and the long-term sustainability of cloud computing resources.

However, AI users can still benefit from public cloud environments and bridge the gap between the freedom of the public cloud and the security and control of an on-premises solution. Indeed, factors such as application availability, performance requirements, and data governance and sovereignty regulations mean organizations should consider on-premises technologies that can deliver the benefits of a cloud operating model while keeping the business entirely in control of the infrastructure supporting these applications. These solutions can provide the best of both worlds: cloud-based economics with on-premises resources, combining their performance and security requirements with the flexibility of embracing a hybrid approach.

High-level Architecture of Lenovo TruScale for AI

The Lenovo TruScale AlaaS offering for AI (Figure 4) is built with this hybrid approach. Lenovo simplifies AI adoption with optimized, ready-to-deploy infrastructure, proven expertise, and pre-validated ISV and partner solutions designed for any size or scale. At the foundation is leadership expertly engineered on-premises high-performance servers and storage for AI ranging from workstations to the edge to the data center to the cloud.

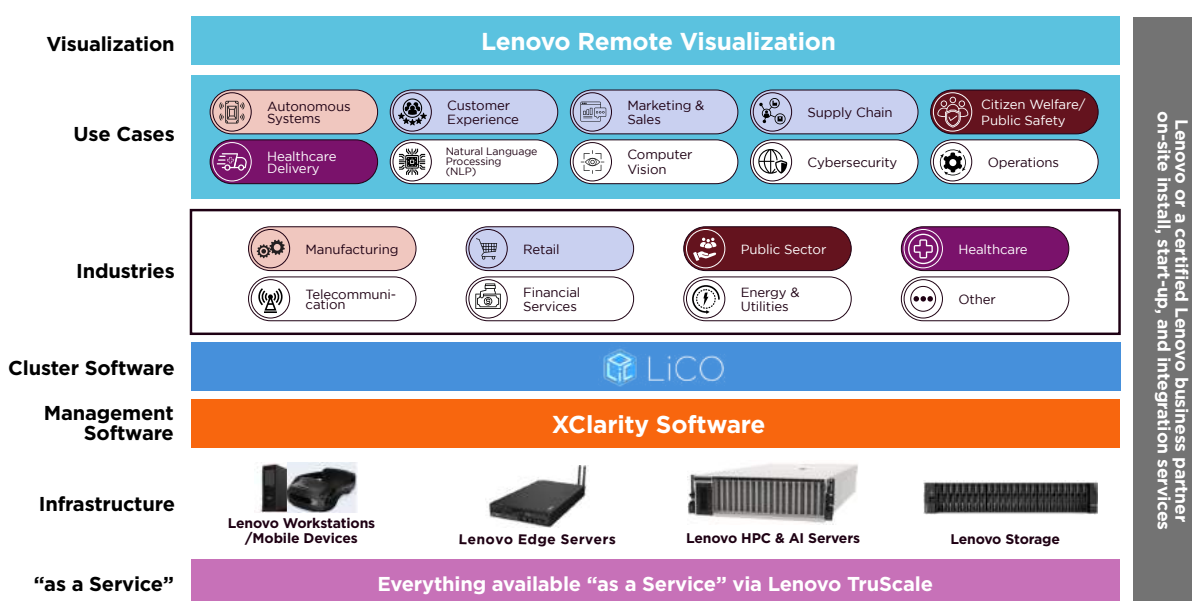


Figure 4: Lenovo AI solutions portfolio available on Lenovo TruScale

Expertly Engineered Lenovo Solutions for AI

Lenovo offers a broad portfolio of mobile devices, workstations, servers, storage, software, and services, many of which are part of Lenovo TruScale (Figure 4). Key components of the portfolio, starting with Infrastructure, include:

- Lenovo Workstations/Mobile Devices:** Lenovo Data Science workstations deliver the high-performance, near-real-time experiences that today's data analytics demand and are ISV-certified, energy-efficient, and highly versatile. These powerful systems are pre-installed with many of the industry's most popular AI frameworks and Data Science/Data Analytics tools and applications. With these systems, you can unleash the potential of your data, allowing business and Data Analytics alike to interactively query, visualize, and explore workstreams with billions of rows and records in mere milliseconds.

Lenovo is implementing AI from the pocket to the cloud with cutting-edge smart devices and

solutions that ensure data science is accessible across all industries in the new hybrid and remote work era. For improved collaboration, new Lenovo ThinkPad commercial laptops include a new Lenovo View application, offering AI-enabled computer vision technology for enhanced video image quality and tools.

Lenovo **ThinkReality XR solutions** help users access immersive simulations. At the same time, purpose-built Lenovo ThinkStation and ThinkPad P Series Data Science workstations deliver the high performance necessary for mission-critical AI model development, data prep, and training tasks.

Generative AI can deliver quality content and solutions to XR at much lower entry barriers and costs than traditional methods, empowering smaller teams or individuals -- even with minimal programming skills -- to create impactful text and higher-quality graphics and art content. Soon, as AI solutions evolve, more applications will arise, such as text to 3D tools and 3D artifacts, helping galvanize the creative process to unleash new levels of creativity.

- **Lenovo Edge Servers:** Deliver purpose-built and secure platforms suitable for compute-intensive and latency-sensitive applications deployed outside traditional data centers. They range from the purpose-built, compact, and secure ThinkEdge SE350 to the AI server for the Edge – ThinkEdge SE450 and the latest ThinkEdge SE360 V2. This new Lenovo ThinkEdge SE360 V2 is the latest purpose-built Edge server that delivers more power to enable AI solutions ranging from computer vision to voice & generative AI. Whether a store, factory, or office, the new server provides extended GPU capabilities supportive of Intel® Flex Series, NVIDIA® L4, and Qualcomm® AI 100 platforms to enable intense workloads at the source of data creation. The ThinkEdge SE360 V2 delivers the highest GPU density in the smallest form factor available. It is at least twice as quiet as competitive products, allowing Edge AI to be deployed in the most remote and rugged sites.
- **Lenovo ThinkSystem Servers:** Highly reliable, scalable, and high-performance servers to significantly accelerate CAE. This [Lenovo portfolio of servers](#) includes **Lenovo Neptune liquid cooling technologies, Lenovo ThinkSystem SR675 V3, Lenovo ThinkSystem SD665 V3, Lenovo ThinkSystem SD665-N V3, Lenovo ThinkSystem SD650-N V2 rack server**; some of these servers also come with **Lenovo Neptune Hybrid Cooling Module**, which quickly dissipates heat in a closed-loop liquid-to-air heat exchanger (L2A), delivering the benefits of liquid cooling without adding plumbing.
- **Lenovo Storage:** [Direct-Attached Storage](#) JBODs and expansion units provide flexible, cost-effective, high-capacity storage and are ideal for space-constrained environments and cost-sensitive customers. [ThinkSystem DE Series All-Flash Array](#) are designed for extreme performance with up to 2.0M IOPS and sub-100 microsecond latency and include industry-leading, enterprise-proven availability, and security features.
- **Lenovo XClarity:** This family of software simplifies and automates the deployment and management of Lenovo infrastructure so clients can focus on their high-value projects.
- **Lenovo Intelligent Computing Orchestration (LiCO):** Reduces the complexity of using a massive HPC cluster and simplifies application deployment, operation, and acceleration.
- **Industries:** Across many industries, fueled by increased global digitalization and demand for IT infrastructure upgrades, Lenovo is simplifying the often-complex implementation of new AI capabilities by delivering AI to the source of data and harnessing its vast network of

best-in-class partners to build next-generation turnkey solutions that enable computing intelligence directly at the edge. This expands its industry-leading AI-ready portfolio of smart devices, infrastructure solutions, and services to help accelerate innovation, enabling the use of generative AI and delivering cognitive decisions at scale across manufacturing, retail, public sector, healthcare, financial, and other industries.

- **Use Cases:** Customers can harness the value of their data, deploying purpose-built AI solutions to transform their business with more predictable outcomes for end-to-end operations, including autonomous driving, customer experience, marketing and sales, supply chain, citizen welfare, public safety, healthcare delivery, security, predictive maintenance, computer vision, natural language processing, virtual assistants for every industry, and more.
- **Lenovo Remote Visualization:** Provides reliable and secure access to graphics-intensive applications anytime, anywhere. Instead of issuing new expensive workstations to all design staff, IT can deploy less expensive enterprise or consumer-class personal computers. In addition, IT departments can maintain security and keep costs down by using remote virtualization hosted in an internal data center or from the cloud. Remote visualization performs intensive graphics operations on a high-end graphics server and generates a 2D pixel version that users can receive quickly. In addition, server-side rendering considerably speeds up the process of using graphics in remote sessions.
- **Lenovo or Certified-partner Services:** Lenovo and an extensive ecosystem of highly specialized AI services partners can deliver the integrated Lenovo stack depicted above. They also provide onsite installation and start-up services to integrate this into a client's work environment, including AI workflows.

Lenovo TruScale includes most of the AI solutions portfolio (Figure 4) as an "as a Service" model and provides all the benefits of public clouds while overcoming their many limitations.

Lenovo TruScale is Affordable and Optimized for AI

High-performance systems are expensive, evolving, and often highly capital-intensive. In addition, users must integrate new technologies and workloads efficiently and seamlessly as technology rapidly evolves, often within the resource, budget, and capital restrictions.

Lenovo TruScale is optimized for today's analytics workloads and designed for future data-intensive, next-generation AI applications. Figure 5 summarizes the key benefits of TruScale for AI.






TruScale for AI	Benefits
 Plug & Play	<ul style="list-style-type: none"> Right-sized AI infrastructure is designed, installed, and configured by Lenovo AI experts Deploy in client's data centers (on-premise) and co-location facility Infrastructure is fully managed and operated by Lenovo AI experts
 Flexible Financial Model	<ul style="list-style-type: none"> Align costs to business outcomes with the Lenovo OpEx solution Minimize up-front capital investment Control costs with Lenovo's predictable pricing model Reduce TCO by helping to eliminate over-provisioning
 Simplified Operations	<ul style="list-style-type: none"> Easily provision AI resources and get support through the Lenovo AI team Centralized management through the Lenovo AI team Get expert help to design, deliver, and operate from a single vendor Incident resolution & support from Lenovo's AI experts, securing the running of operations and uptime
 Latest Technology	<ul style="list-style-type: none"> Access to the latest technology to help run client's environment No need for capital investment: avoid long procurement approvals with Lenovo's OpEx model Technology is refreshed as it becomes obsolete, Lenovo will manage the upgrades and replacements
 Customized Solution	<ul style="list-style-type: none"> Solution customized and optimized to a client's specific AI computational needs

Figure 5: Benefits of the Lenovo TruScale solution for AI

TruScale is a new way of procuring IT resources via a consumption-based subscription model. With this new model, customers never take capital ownership of the hardware or other IT assets. They pay for what they use each month only for capacity when their workloads are actively running as part of their operating expenses with no minimum capacity commitment. Lenovo TruScale includes hardware installation, deployment, management, maintenance, and removal.

Lenovo's unique metering solution remains outside the customer's data plane – providing the advantages of cloud-like economics with the security of on-premises hardware. Additionally, capacity can be scaled up or down to accommodate business needs, ensuring IT infrastructure is always right-sized. In one bill, monthly pricing structures are simple and all-inclusive of associated services (maintenance, support, remote monitoring, and system health).

Worldwide, many clients benefit from the unique value provided by the Lenovo AI Infrastructure.

However, these processes are powered by ML and assume that all data consumed is the same within the process, meaning that data is considered to be independent and identically distributed (iid), but that is not the case. One example is when data is assumed to be iid but is correlated in different contexts. This is called a domain shift, which is extremely common in ML.

To overcome these limitations, Lenovo collaborated with the Barcelona Supercomputing Centre (BSC) and researched how autonomous driving data, known as SODA 10M, are consumed and processed. The concept of Continuous Learning (CL) keeps an ML system learning post-deployment, e.g., within an autonomous vehicle. Traditionally, this is done by retaining all learned data and then retraining the system frequently. However, due to various guardrails, this can pose problems around data privacy, storage, or compute restrictions. These limitations create a problem known as “catastrophic forgetting,” in which a neural network cannot correlate what it previously learned and the new data it was provided.

Client Case Studies on Lenovo AI Infrastructure

Globally, Lenovo's high-performance infrastructure has been deployed at leading organizations – many of whom are doing pioneering work in AI to drive digital transformation and get better outcomes.

Autonomous Driverless Cars⁴: Traditional autonomous vehicles use ML to drive from a starting point to their destination, enabled by various technologies, including active steering, adaptive control systems, GPS, lasers, LIDAR, radar, and sensors.



The research was done on a Lenovo ThinkSystem SR650 V2 server. A task detector that could operate unsupervised was developed and did not have to be informed that any data was changing or carrying a new label. With this development, the ML within the vehicle did not need to update but could learn and run within the context switch it was experiencing first-hand.

Autonomous vehicles are progressing to advance and become a regular occurrence in day-to-day life for society. Lenovo, alongside its partners and fellow scientists, will continue to help advance this technology and continue to drive smarter technology for all.

Retail⁵: E-commerce and online shopping have strongly affected the retail industry in the last decade, which has transformed post-COVID-19, with thousands of retail store closings and scores of national chains declaring bankruptcy. One way to overcome these challenges is to deploy cutting-edge technologies like AI, augmented reality (AR), virtual reality (VR), edge computing, and ML.

AI can help retailers detect fraudulent activities such as fake reviews, credit card fraud, and account takeovers and take appropriate action to prevent them. It can also help retailers optimize their inventory management by analyzing customer data, including past purchases and browsing history, to make personalized recommendations for products and services they are likely interested in and ensure they are in stock when customers arrive. One of the main reasons a well-known major national chain went bankrupt in the last decade was that its supply chain logistics was highly inefficient—the inventory database was not integrated with its sales database. Thus, items sold were never replenished, disappointing customers who could not purchase what they wanted. A well-designed and implemented supply chain management driven by AI can improve the customer experience and increase sales.



AR/VR in retail can also dramatically improve how customers engage with brands and boost involvement. AR is being used to help guide customers through a store to locate items faster.

It also can be used to display products virtually. Ikea used this technology so customers could **virtually place products** in their homes to see how they looked. And with VR, customers can try on items or tour a store from the comfort of their couch. Since the start of the pandemic, virtual samplings have **increased 32%**.

“Everseen’s Visual AI™, together with Lenovo Edge AI servers and NVIDIA® Accelerated Computing AI Platform, are helping us shape a smarter retail experience that reduces friction for our customers and makes our operations safer and more efficient.”
Chris McCarrick, Senior Manager of Asset Protection Solutions & Technology, Kroger

Cincinnati, Ohio-based Kroger Co. is the nation’s largest grocery chain, with annual revenue of about US\$138 billion. It operates supermarkets, pharmacies, fuel centers, and multi-department stores throughout the U.S. To improve customer experience, the chain has been expanding the use of self-service checkouts (SCOs) nationwide. Although SCOs simplify and speed up the checkout process, they still have some sticking points. For example, while scanning items, customers often cannot get the scanner to recognize a barcode, or a customer might neglect to scan an item either intentionally or accidentally. Errors like these impact both Kroger’s bottom line and customer experience.

To address these issues, Kroger deployed Visual AI™ Application from Everseen on high-performance Lenovo Edge AI servers, powered by NVIDIA® Accelerated Computing AI Platform, to analyze the video footage from checkout kiosks in real time to recognize regular processes and intelligently step in whenever something is amiss. For example, if a shopper fails to scan a particular item successfully, the SCO system flags the error on the screen and prompts the customer to self-correct. If the customer cannot resolve the issue, the system alerts a store associate via a mobile device to intervene and rescan the item. Since deploying this solution, over 75% of the time, customers can resolve scanning errors by themselves without sales associates’ intervention, making the process a little easier.

In summary, ML models for retailers help them quickly review and convert a large amount of complex data into insights that can be used to:

- Improve customer shopping experience.
- Accurate forecasting of upcoming needs.
- Improve inventory control.
- Identify consumer needs through appropriate segmentation.
- Make product offerings more unique.
- Decide on the optimum prices to increase sales.

Technology and Innovation Park⁶: Lenovo recently signed an agreement with Dubai, UAE-based **AI Hathboor Bikal.ai** to provide public and private sector customers with cutting-edge Lenovo TruScale High-Performance Computing (HPC) clusters to support advanced digital transformation projects which focus on developing innovation in the UAE.

Lenovo will also provide systems and expertise to help AI Hathboor Bikal.ai to build and operate a new sustainable data center at **Sharjah Research Technology and Innovation Park** (SRTIP).



Lenovo HPC clusters will play a crucial role in citizen welfare, public safety, and health by providing powerful computational resources to AI Hathboor Bikal.ai by addressing complex challenges and improving decision-making processes. Here are some applications of HPC clusters in these fields.

“We have huge ambitions for Sharjah Research and Technology Innovation Park, and this partnership with a global leader like Lenovo, will set the path towards developing tailored infrastructure solutions that will solve both public and private organisational challenges. We look forward to leveraging Lenovo’s experience in the technology sector and enhancing our digital capabilities while also welcoming new AI-powered and data-enabled research and technologies.”

Raj Sandhu, GM AI Hathboor Bikal.ai

Disaster Management and Emergency Response: HPC clusters will help simulate and predict the impact of natural disasters, such as hurricanes, earthquakes, or floods, enabling authorities to develop effective evacuation plans, allocate resources efficiently, and assess potential risks. Real-time data analysis from various sources, including weather sensors, satellite imagery, sensor networks, and social media feeds, can be processed rapidly on HPC clusters, and advanced analytics capabilities facilitate predictive modeling and machine learning algorithms to support emergency response efforts.

Public Health Analytics and Surveillance: HPC clusters can handle extensive data processing tasks involved in public-health analytics and surveillance. They enable real-time monitoring and analysis of health-related data streams, including CCTV, electronic health records, medical imaging data, sensor data from wearable devices, and social media feeds. Typical applications include Drug Discovery and Development and Medical Imaging and Diagnostics. These clusters can help identify disease trends, detect outbreaks, and support early warning systems.

“AI Hathboor sees this collaboration as a way to give the UAE a technological advantage to the innovation ecosystem, create jobs and opportunities for our youth, and develop tech for solving regional and global challenges. Lenovo, Nvidia® and Intel® with SRTIP gives us a joint capability in tech transfer and we look forward with working with our universities, businesses, and government”

Mohammed Al Hathboor, Management Committee Member of AI Hathboor Bikal.ai

Overall, Lenovo TruScale HPC clusters will provide the computational power to handle large-scale data analysis, simulation, and modeling tasks in citizen welfare, public safety, and health. They enable researchers, scientists, and public health professionals to gain valuable insights, make informed decisions, and improve public safety and health outcomes.

Get Started with Lenovo TruScale with Dedicated Support

Organizations deploying AI in production cannot afford performance problems, delays, or downtime. Therefore, support must be proactive, carried out by technical specialists who work closely with the customer and deeply understand their environment.

Lenovo

As part of their Lenovo TruScale for AI contract, businesses receive a dedicated technical account manager or system admin to serve as their single point of contact. Whether onsite, working remotely, or a mixture of both, support professionals can quickly pinpoint and resolve any issues and ensure the AI environment runs optimally 24/7.

However, Lenovo goes way beyond specialized technical support. TruScale's end-to-end service for AI includes initial consultation, analysis, and configuring the right environment through ongoing cooling assessment and maintenance services to billing and administration.

The Lenovo TruScale Advantage for AI

Lenovo TruScale helps organizations of all sizes maximize their return on investment in their AI infrastructure solutions and accelerate time to value to drive innovation.

It also gives the flexibility to rapidly add provisioned resources without being hindered by procurement delays or supply chain disruptions – something vital today.

Consuming your existing high-performance infrastructure for AI is more transparent and affordable (you get charged only for what you use with a detailed report of your metered bill) than building new traditional on-premises solutions without the limitations of public clouds. Furthermore, as every organization has unique needs and requirements, working with Lenovo will help achieve a solution tailored to their evolving workload, workflow, and workforce needs.

Lenovo
TruScale

Start planning your cloud
journey with Lenovo

Arrange a discussion with one of our experts now.
truscale@lenovo.com

Lenovo.com

1. McKinsey, "The state of AI in 2022—and a half decade in review", December 2022
2. Wall Street Loves AI: Goldman Expects A Shocking \$7 Trillion Boost In Global GDP And Raises Targets For MSFT, CRM, ADBE - Adobe (NASDAQ:ADBE), Amazon.com (NASDAQ:AMZN) - Benzinga
3. Challenges in Deploying Machine Learning: A Survey of Case Studies, 2011.09926v2.pdf (arxiv.org), Jan 2021
4. <https://lenovopress.lenovo.com/lp1716-autonomous-vehicles-use-new-ai-algorithm>
5. <https://news.lenovo.com/shaping-krogers-self-checkout-of-tomorrow-ai-edge-computing/>
6. New HPC cluster Enables Lenovo & AI Hathboor Bikal.ai to Drive Citizen Safety & Health from Sustainable Data Center - Lenovo StoryHub

Lenovo reserves the right to alter product offerings and specifications, at any time, without notice. Lenovo makes every effort to ensure accuracy of information but is not liable or responsible for any editorial, photographic, or typographic errors. Images are for illustration purposes only. For full Lenovo product, service, and warranty specifications, visit www.lenovo.com. Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo. Other company, product and service names may be trademarks or service marks of others.
© Lenovo 2023. All rights reserved