GIGABYTE™

NVIDIA

# HOW TO MONETISE AI FOR BETTER BUSINESS OUTCOMES

IDG

# AI in the modern business landscape

Increasing connectivity has significantly shifted the world of business, producing opportunities and challenges in equal parts. Labour and capital now move freely across borders as with the prospect of growth, and ideas and innovation are sped up thanks to deeper collaboration. Yet at the same time, competition for talent, resources, and customers are heightened, with accelerated disruption intensifying volatility across almost every industry.

This, combined with the cornucopia of global online marketplaces and hyperconnectivity at the fingertips, resulted in the modern, savvier, and more complex generation of consumers—one who looks beyond traditional price and quality factors. Today's consumers also take into consideration convenience, customisation and control, brand advocacy, always-on service and support, and digital trust and security into their decision process.

This growing complexity of the business landscape and consumer base has led to more enterprises turning to artificial intelligence (AI) to unlock insights and stay ahead of consumers. In Asia, organisations are starting to see AI as one of the core drivers to accelerate digital transformation, develop or grow digital resilience, and sharpen their competitive edge in the global playing field.

In a Nov 2020 IDC study, 87% of CXOs said that becoming a more 'intelligent enterprise' is among their top goals within the next five years, while IDG's annual Digital Business research in Aug 2021 showed respondent organisations are expecting to invest an average of US$16.5 million on digital initiatives over the next 12 months, with 38% (the highest percentage) saying they anticipate spending even more on AI and Machine Learning (ML) tools.

While organisations are employing AI to achieve a wide variety of business outcomes, their efforts broadly fall into three main use cases:

### Delivering delightful experiences

AI is used to improve the customer experience of existing products and services. This includes providing recommendations based on customers' preferences for a personalised experience or enhancing functions with digital assistants.

### Improving processes and technological efficiency

AI is deployed to enhance business workflow, whether in offices or across shop floors. AI gets the job done by conducting proactive surveillance, drafting predictive and preventive maintenance schedules, and streamlining inventory management.

### Enhancing end-user engagement

AI provides the bridge between growing demand for deeper customer engagement and growing labour costs. The next generation of customer support and engagement will employ increasingly intelligent, sophisticated chatbots and interactive automated voice responses as a standard baseline.

With wide applications and compelling use cases, AI adoption continues to increase. IDC's Feb 2021 AI Tracker reported that the global AI market is projected to grow 16.4% year-on-year in 2021 to a market size of US$327.5 billion, and the market will break the US$500 billion mark by 2024. A separate IDC report in Apr 2021 further showed that 89% of organisations interviewed in APAC (excluding Japan) had already started on their journeys to harnessing AI.

AI's importance in the modern business landscape is on the rise, and it is fast becoming an imperative for CIOs to capitalise on AI's potential.

# Barriers to AI adoption

Despite the promise of AI being widely recognised by enterprises, most organisations in Asia continue to **lack an organisation-wide strategy** that could maximise AI's potential. Of the 89% of organisations that had already made a start with AI adoption, 52% of those in Asia admitted that their current investments in AI had ended up being used in siloes, or only by select groups or for isolated projects.

Besides the absence of a comprehensive AI strategy and roadmap, many organisations also **underestimate how bandwidth intensive it can be to support AI workloads**. These workloads can include the software and platforms needed to build AI capabilities, as well as AI applications based on ML and deep learning from unstructured data and information, and make heavy use of memory and parallel computing to carry out bulk floating-point operations.

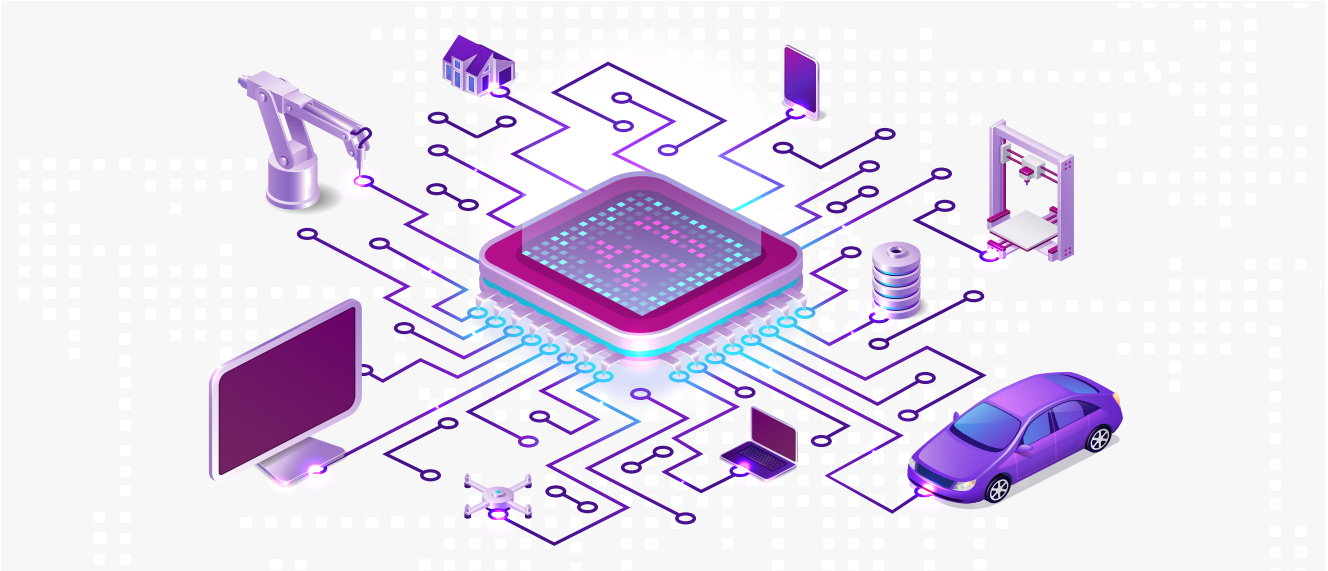
This need for high-end computing performance continues throughout the entire AI lifecycle: from data collection and preparation, to building and training an AI model and then inferencing it with data, to scaling the AI model and operationalising it across the enterprise.

**Disparate existing infrastructure** also poses a challenge to AI integration. Enterprises may choose to build an end-to-end AI lifecycle across multiple different products, and then attempt to integrate with existing infrastructure to support the AI workloads. A 2020 Gartner survey highlighted that only 53% of AI projects make it from pilot to production phase, and the complexity of achieving AI solution integration with existing infrastructure was cited among the top three barriers to successful implementation.

For enterprises that do successfully achieve proof-of-concept by building and running an AI model, deploying it throughout the enterprise requires **effective scaling**. AI applications will need fast central processing and fast networking support as they work through the vast data sets generated at enterprise level.

Resources should also be catered to ensure **management of the AI model**—embedding the model to new and existing enterprise applications and processes, integration with existing enterprise infrastructure, monitoring and iterative improvement, as well as continual updating of data, governance, and compliance requirements—and **system availability**.

Before embarking on the AI journey, organisations must make a realistic assessment of the costs in time, finance, and manpower that come with building and deploying an AI model and should understand that rapid development and deployment is not easily achievable, especially if starting from scratch.



# Overcoming the AI workload challenge

Recognising these challenges, vendors are developing solutions for enterprises to effectively power, support and manage their AI workloads throughout the AI lifecycle.

Each enterprise will have specific requirements, depending on the state of their AI journey, the speed at which they aim to progress and the desired end-state to be achieved. In general, enterprises should be mindful of what they need when selecting the right AI lifecycle solution, in terms of:

### Delivering the Right Level of Performance

AI workloads must be sufficiently supported to handle the large data volumes needed for initial training and inferencing of the AI model.

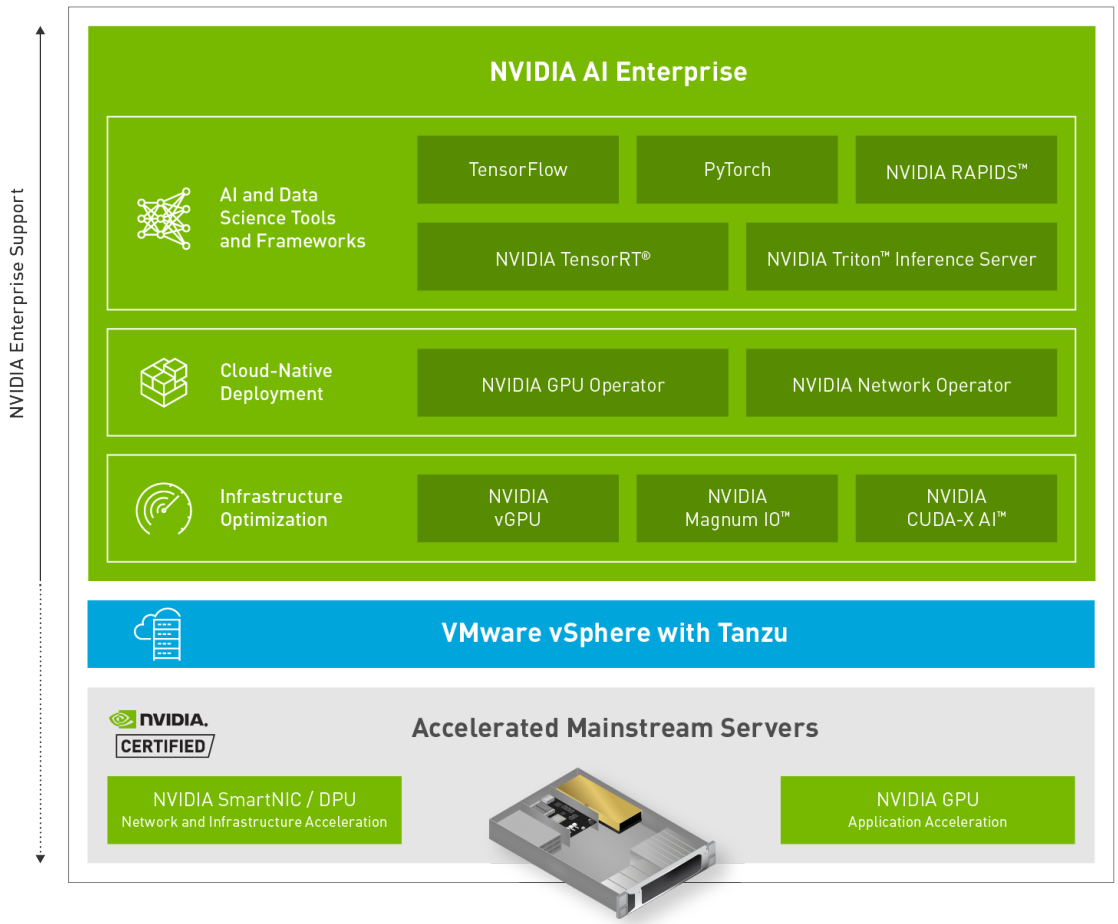
### Delivering Scalability

Beyond proof-of-concept, the AI platform must be able to bring the project from testing to enterprise-wide deployment, both in terms of hardware as well as software.

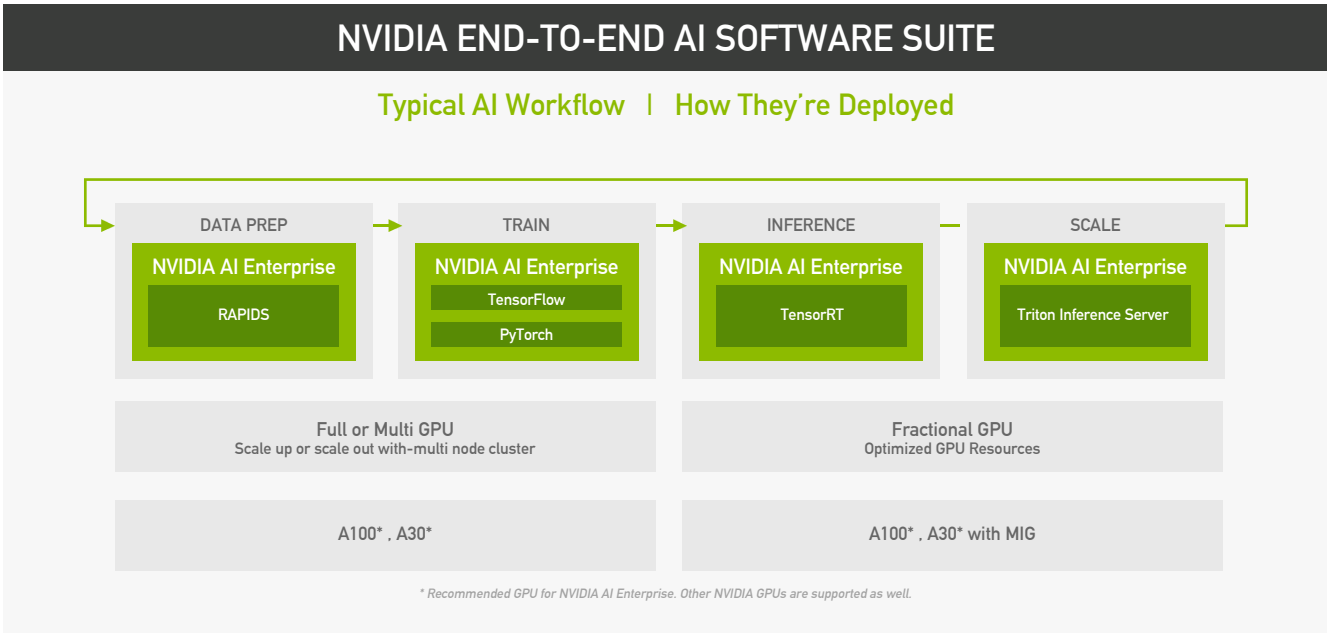
### Delivering Simplified Management

An AI solution must be able to bring together and ensure interoperability across disparate products, as well as integration with existing enterprise infrastructure.

NVIDIA AI Enterprise with VMware on GIGABYTE servers offers a full stack, end-to-end platform that is optimised for supporting AI workloads, delivering a suite of cloud-native AI and data science applications and frameworks. It includes key enabling technologies from NVIDIA for rapid deployment, management, and scaling of AI workloads in the modern hybrid cloud.



The NVIDIA End-to-End AI Software Suite can be deployed across the entire AI workflow, while optimising GPU resources at the various stages of the AI lifecycle.




By employing NVIDIA AI Enterprise with VMware on GIGABYTE servers, enterprises can expect a range of benefits including:

- **Rapid deployment,** with the confidence of jointly certified NVIDIA and VMware solutions on the best combination of hardware and software, optimised to support various AI workloads
- **Simplified management and seamless deployment** in existing enterprise infrastructure environments
- **Near bare-metal performance** for AI training and inference, with enterprise-grade security, manageability, and efficiency features
- **High availability** for AI workloads with simplified infrastructure maintenance (consolidation, expansion, upgrades) for easy scaling up or down
- **Ability to store and process vast datasets,** as well as massive parallel computing capabilities needed to train algorithms powered by GIGABYTE's storage and high-density servers
- **Customisability to address specific requirements,** which can be validated as NVIDIA-Certified Systems with GIGABYTE's wide range of servers

The increasingly complex, hyperconnected landscape and sophisticated consumer base is leading many organisations to turn to AI for better insights, efficiency, and next-generation customer engagement. But building and deploying an effective AI can be demanding, and organisations must be prepared to invest the manpower, cost, and time to see the journey through. Enterprises will need to mitigate the risks arising from building an AI solution across disparate products and existing infrastructure, invest in high performance computing capabilities to sustain AI workloads, and manage costs and resources to ensure effective scaling for enterprise-wide deployment.

Fortunately, solutions such as the NVIDIA AI Enterprise with VMware, deployed on GIGABYTE servers—combining leading-edge software and robust hardware—offer assurance and confidence in supporting AI workloads from data collection to scaled-up deployment, leveraging end-to-end, cloud-native applications that are optimised for NVIDIA-Certified Systems. With assured security, performance, and scalability , the right solution can provide the cornerstone to a successful enterprise-wide AI strategy for the future.