Realizing the Potential of AI in the Retail Industry

Overcome challenges and deliver on the full promise of artificial intelligence with the AI-Ready Enterprise Platform, featuring the NVIDIA AI Enterprise software suite in VMware virtualized environments

The Power of AI vs. the Obstacles of Infrastructure

Retail is now the dominant industry in AI investment.

In 2021, retailers spent about $12 billion on AI. The retail and banking industries, the two largest adopters of AI, will account for nearly 28 percent of all AI expenditures in the U.S. by 2025, and global retail spending will continue to outpace that of banking over the next five years.

These massive AI investments fall into three main categories for retailers: **omnichannel management, intelligent supply chain,** and **intelligent stores**. In sum, AI is reshaping every facet of the industry, from computer vision in distribution centers to autonomous checkout.

But spending doesn't automatically equate to success. There are common concerns for retailers deploying AI, including scaling from proof-of-concept to implementation on legacy infrastructure. In fact, almost half of all AI projects never make it to production[1], hindered by familiar challenges.

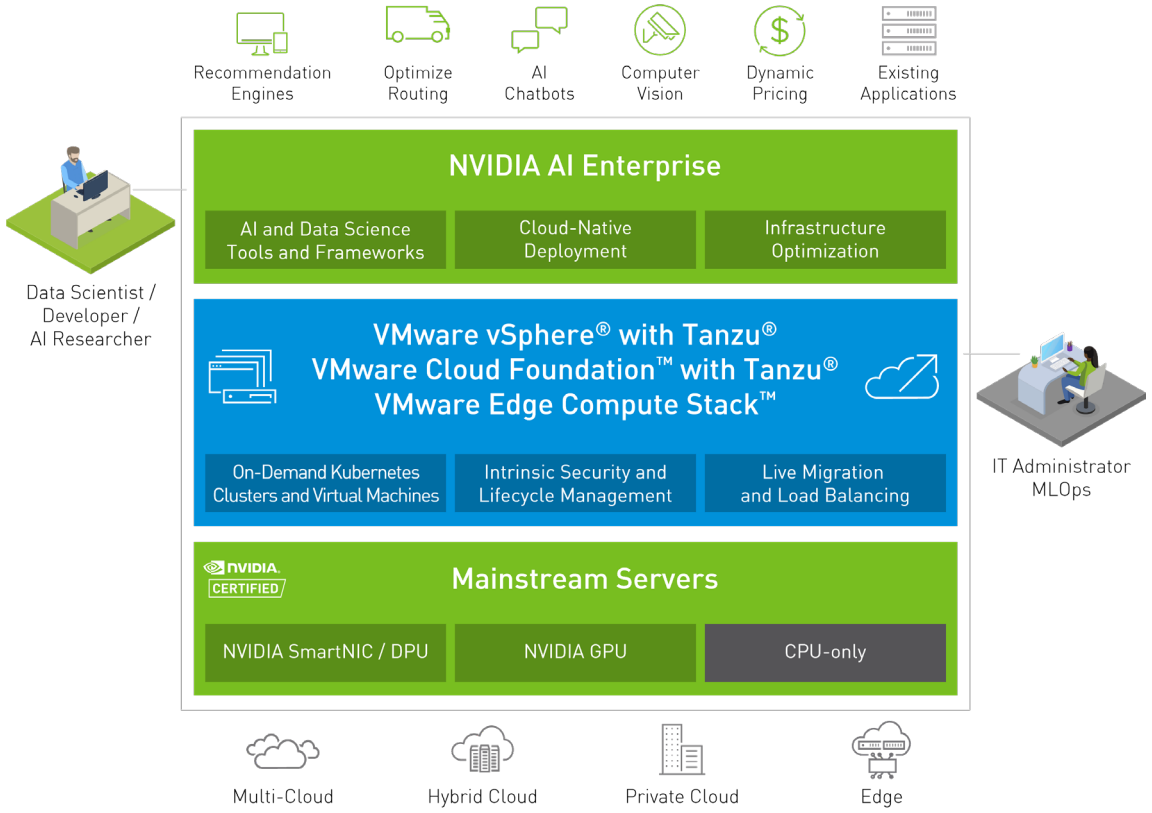
- **Incompatible infrastructure:** The complexity of deploying an end-to-end AI solution onto existing infrastructure is a primary barrier, especially in distributed retail environments, where AI workloads must perform at edge locations.
- **Data silos and shadow IT:** Investment often happens at a project or team level, on customized bare-metal infrastructure to achieve high performance, which creates pockets of AI outside central IT control.
- **Trouble scaling and lagging performance:** Bare metal makes scaling AI more difficult; however, data scientists without high-performing systems waste weeks training models.
- **Cost control:** Disparate IT and unused infrastructure raise overhead and cost per workload.

> **Everything you can think of in almost every part of retail is being powered by AI.**
>
> **– Jeremy King**
> *Senior vice president of engineering at Pinterest, in The Wall Street Journal*

1. Gartner "P-19019 AI in Organizations," Claudia Ramos, Erick Brethenoux, 2020

AI-Ready Enterprise Platform: AI Simplified

The AI-Ready Enterprise Platform, a joint solution from VMware and NVIDIA, allows retailers to deploy rapidly onto their existing IT infrastructure, including at edge locations common among brick-and-mortar chains. It's interoperable from the hardware to the application layer, offering shared infrastructure for both AI workloads and core retail apps.



Features and Benefits

Simplicity and scalability: NVIDIA-Certified Systems™ and VMware vSphere® with Tanzu®, the industry's leading virtualization platform, offer ease of deployment and scaling for AI workloads while providing a common system for both virtual machines and containers.

Accelerated AI adoption: State-of-the-art AI frameworks and tools from NVIDIA enable data scientists and AI practitioners to deliver business insights faster.

Manageability: The platform eliminates data silos and shadow IT by merging AI with mainstream data centers, most of which run on VMware systems, allowing companies to extend existing vSphere management processes and skills into the AI space.

Reduced cost without compromising performance: Virtual GPUs and accelerators can be shared, driving down total cost of ownership, and infrastructure optimizations enable the bare-metal performance and multi-node scale required for big data sets and training large AI models.

Data-Driven Omnichannel Management Fuels Sales

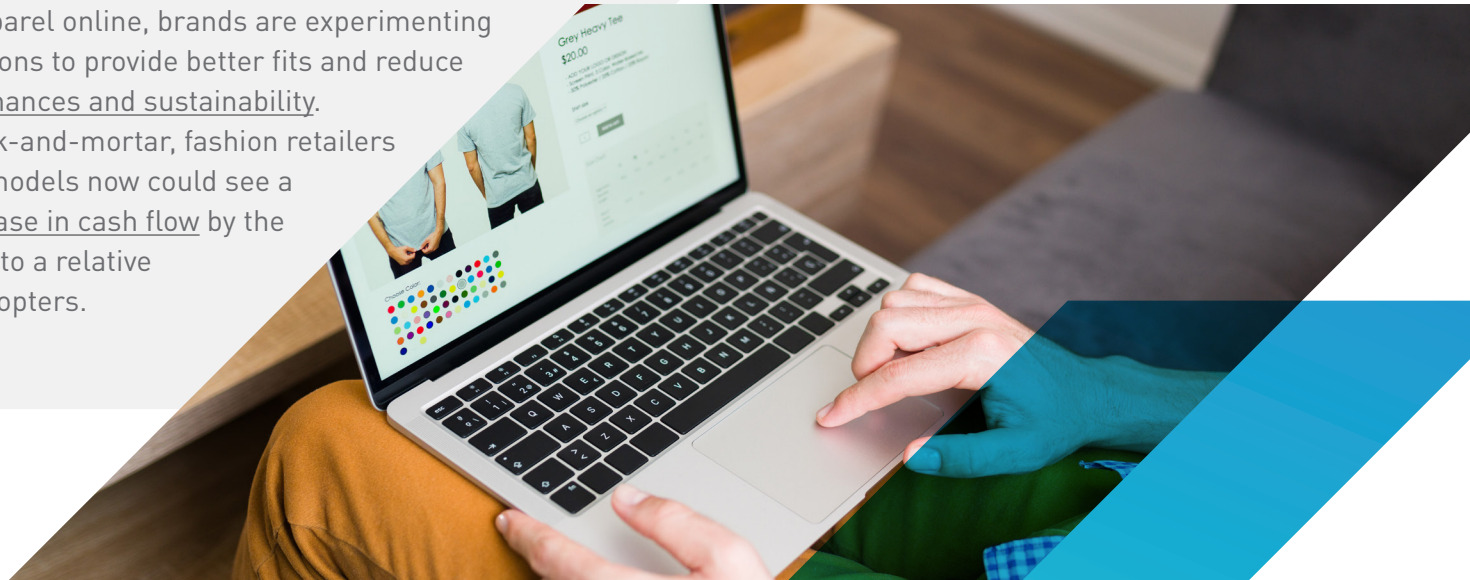
Within omnichannel management, NVIDIA and VMware's enterprise solution helps retailers create 360-degree views of customers to increase personalization and revenue.

Augmented customer service agents and product recommendations account for nearly 40 percent of industry investment on AI because they are proven to increase sales and engagement. To enhance customer service, AI chatbots and intelligent voice assistants use natural language processing (NLP) to connect across an expanding number of touchpoints—in-store, online, and in-app. Meanwhile, product recommendations already account for as much as 30 percent of revenue on large commercial internet platforms, and more accurate recommendation engines—backed by NVIDIA GPU-powered machine learning (ML) algorithms—can improve e-commerce revenue by 60 percent.

AI apps are also transforming retail. Some gamify shopping to better understand consumer preferences, while others provide visual search capabilities, allowing users to search or receive recommendations based on an image's characteristics.

Industry Spotlight: Online Shopping

As more people purchase apparel online, brands are experimenting with AI-generated virtual try-ons to provide better fits and reduce returns, thereby improving finances and sustainability. Across e-commerce and brick-and-mortar, fashion retailers embedding AI into business models now could see a 118 percent cumulative increase in cash flow by the end of the decade, compared to a relative 23 percent decline for late adopters.



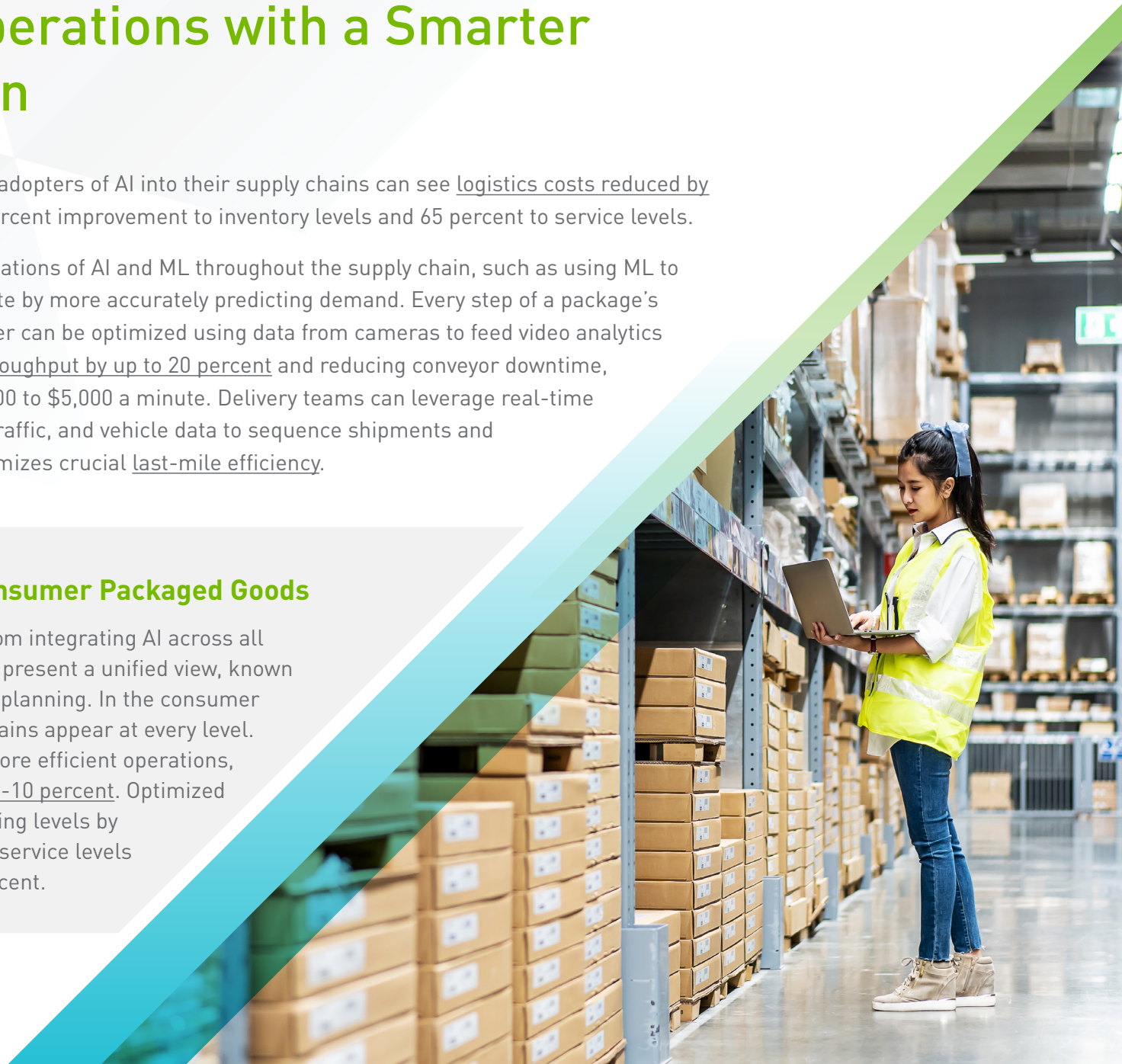
Optimize Operations with a Smarter Supply Chain

Regardless of industry, early adopters of AI into their supply chains can see logistics costs reduced by 15 percent, as well as a 35 percent improvement to inventory levels and 65 percent to service levels.

For retailers, there are applications of AI and ML throughout the supply chain, such as using ML to reduce overstocking and waste by more accurately predicting demand. Every step of a package's journey at a distribution center can be optimized using data from cameras to feed video analytics applications, accelerating throughput by up to 20 percent and reducing conveyor downtime, which can cost retailers $3,000 to $5,000 a minute. Delivery teams can leverage real-time geographic, environmental, traffic, and vehicle data to sequence shipments and optimize routing, which maximizes crucial last-mile efficiency.

Industry Spotlight: Consumer Packaged Goods

The biggest benefits come from integrating AI across all phases of the supply chain to present a unified view, known as autonomous supply-chain planning. In the consumer packaged goods sector, the gains appear at every level. Improved planning creates more efficient operations, reducing supply chain costs 5-10 percent. Optimized inventory can decrease stocking levels by 20-30 percent, and improved service levels can increase revenue 3-4 percent.



Build a Better Brick-and-Mortar

Deploying AI technology in-store at edge locations is critical in the retail industry, where real-time insights are required. The accelerated inference speed provided by NVIDIA GPUs is essential for creating intelligent stores. For example, global retailers lose over $100 billion annually to shrinkage, but intelligent video analytics used in asset-protection solutions can improve loss-prevention efforts. That computer-vision technology uses AI inference for real-time object detection with existing cameras and point-of-sale (POS) systems.

Cameras inside intelligent stores can also work alongside sensors and POS systems to create robust in-store analytics about customer behavior and preferences. Other AI uses include cameras and robots that monitor inventory levels and alert associates to help prevent stockouts, boosting sales. And of course, AI is ubiquitous in autonomous stores that can provide increased margins from unstaffed, fully automated shopping experiences.

NVIDIA has pre-trained models and ready-made applications from its ISV partners for the full spectrum of retail uses, including dynamic pricing and promotion, inventory management, and store analytics.

Industry Spotlight: Grocery

Grocery stores are particularly well-suited to harness AI and ML because of "high purchase frequency, rich customer data, and an emphasis on efficiency," according to a McKinsey report. More accurate demand forecasting can inform store-specific SKU selections that are tailored to hyperlocal consumer needs, resulting in sales increases of 2-4 percent. Personalized promotions can provide increases of 4-8 percent in sales and 2-3 percent in net income and EBIT.

Ready to Get Started?

Experience what NVIDIA AI Enterprise can do with immediate trial access to NVIDIA LaunchPad, a free program that allows enterprises the chance to try end-to-end AI solution workflows for testing, prototyping, and deployment. It runs on private accelerated computing infrastructure in VMware vSphere environments, offering familiar toolsets and systems, and includes these hands-on labs for data scientists, AI practitioners, and IT staff:

AI Development and Data Science

- Train and Deploy an AI Support Chatbot
- Train AI Models for Image Classification of Online Products
- Accelerate Data Processing, Tokenization, and Train an AI Model to Perform Sentiment Analysis
- Accelerate Data Processing and Train an AI Model to Predict Prices
- Scale Data Science with Domino MLOps Platform

Infrastructure Management

- Configure, Optimize, and Orchestrate Resources for AI and Data Science Workloads with VMware Tanzu
- Configure and Optimize VMware vSphere for AI and Data Science Workloads

Learn more about the platform by visiting the VMware and NVIDIA AI-Ready Enterprise site and the retail solution brief.

NVIDIA | vmware