ERICSSON

In partnership with

The network blueprint for AI in retail

How to avoid performance, security, and uptime issues as AI scales

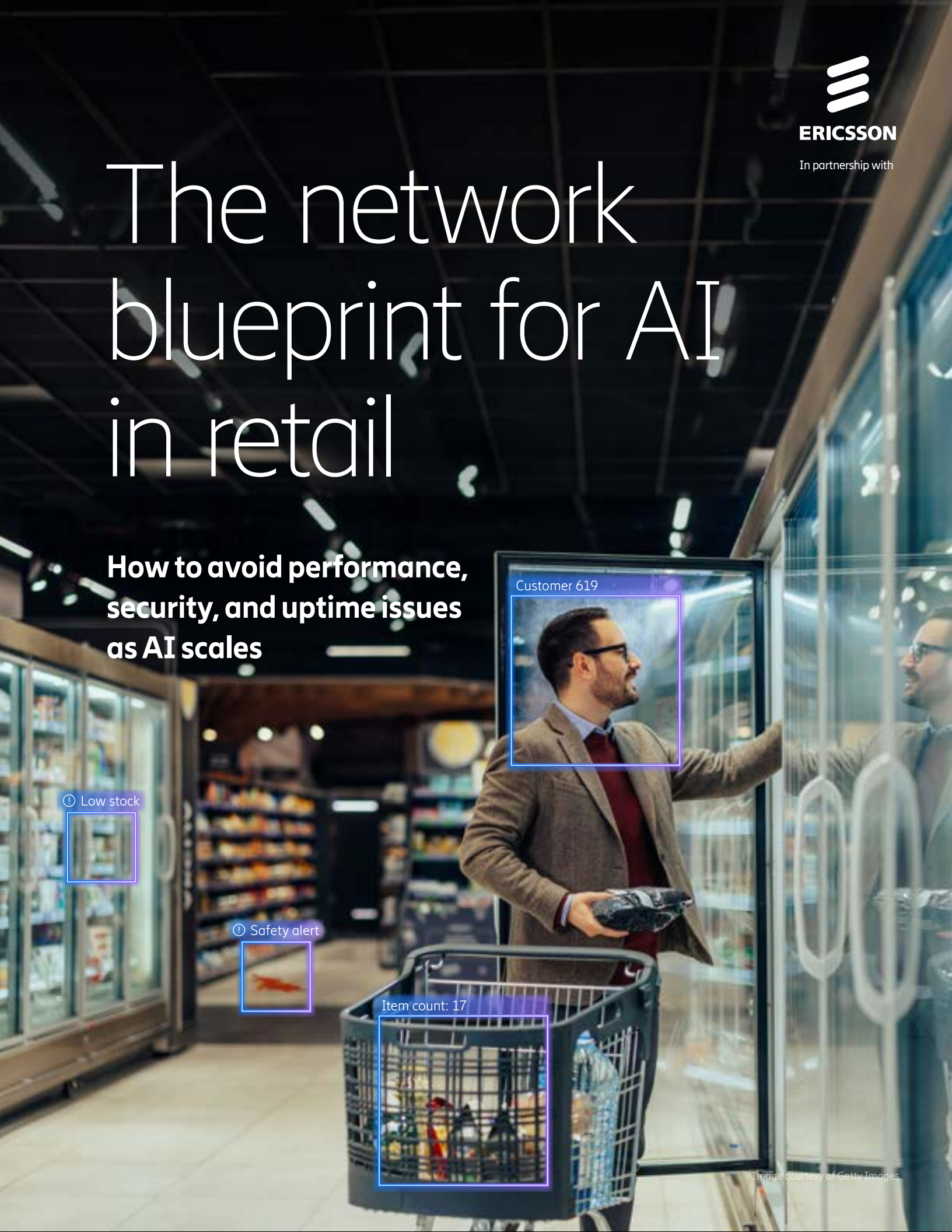


Image courtesy of Getty Images

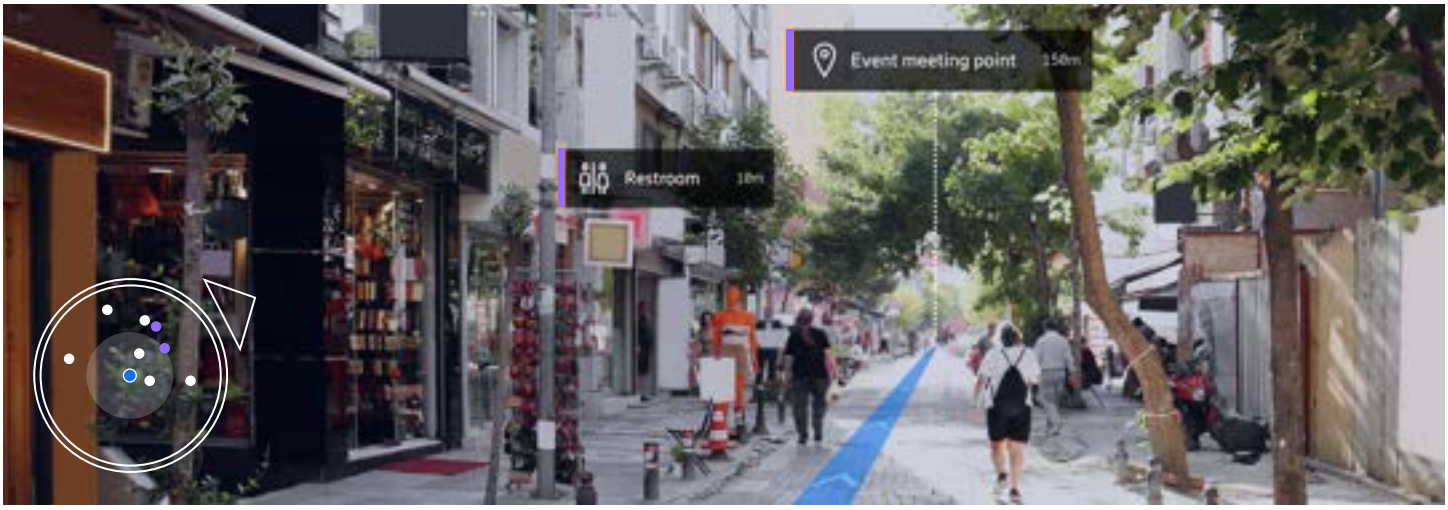


Table of contents

03 The forces pushing retail toward intelligent ecosystems

04 AI architectures and what they demand from retail networks

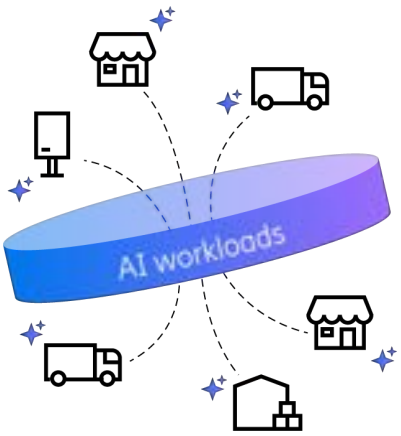
05 Why AI-based applications are best executed at the edge

06 5G and SD-WAN as the foundation for AI-driven retail

07 Best practices for integrating AI in retail

08 Takeaway: What you need to know about AI in retail

In retail, AI can personalize a shopper's journey in real time, help associates find answers faster, flag shrinkage before it becomes loss, and keep inventory flowing across a 24/7 supply chain. But the rise of AI has clear implications for retail IT and network administrators.



AI workloads don't exist in a single location. They span a distributed footprint that can include multiple locations, pop-ups, kiosks, warehouses, and delivery fleets. Such use cases only scale when the network delivers consistent performance, availability, and security across locations.

Two shifts have raised the urgency to modernize the network foundation for AI. First, AI capabilities have become easier to access, moving from background analytics to tools that support employees and customers in day-to-day workflows. Second, 5G has advanced with standalone capabilities and features such as network slicing, which can help prioritize essential traffic and support app performance when the network is congested. Together, these shifts increase the network requirements for retailers that want to capture AI value at scale.

The forces pushing retail toward intelligent ecosystems

The connected store retail era focused on bringing locations online and standardizing core services such as POS, guest Wi-Fi, basic monitoring, and click-and-collect or buy online, pick up in store (BOPIS). The next era favors an intelligent ecosystem where decision-making is distributed, data moves continuously, and frontline operations embrace automation. Three primary challenges are driving this evolution:

Consumer expectations continue to rise.

Once shoppers experience conveniences such as inventory visibility, pickup options, and personalized recommendations, they begin to expect that level of ease across all retail interactions. Even minor digital friction, such as poor in-store connectivity, can disrupt the shopping experience, especially in large-format stores where cellular signal and guest Wi-Fi are often difficult to access or unreliable.

Retail's footprint creates networking problems.

Retail IT teams are expected to deliver predictable application experiences and consistent policy enforcement but often must do so while grappling with different last-mile providers, inconsistent cellular coverage, and varying on-site constraints. The complexity shows up in the need to manage day-to-day realities without relying on on-site support and third-party truck rolls.

IT teams are being asked to do more with less.

Staffing and budgets have tightened while expectations for uptime and digital experience have increased. Vendor sprawl and fragmented management tools add operational drag, especially when diagnosing issues requires switching between multiple systems and logins.

These challenges manifest as data silos, time-sensitive application behavior, and an increasingly broad attack surface. While many retailers see AI as a silver bullet to address their operational shortcomings, AI — if deployed haphazardly — can amplify each constraint by increasing data movement and automation, turning policy gaps, routing instability, and segmentation weaknesses into business risks.

AI architectures and what they demand from retail networks

AI can be grouped into three architectures: predictive, agentic, and generative. Each changes traffic patterns and network requirements differently. Predictive AI increases data synchronization, while agentic AI demands low-latency control and strong segmentation. Generative AI often involves sensitive data, so it needs consistent controls across all locations.

Predictive AI

Anticipating demand and optimizing operations

Predictive AI (traditional machine learning) delivers value in retail by turning large volumes of historical and time-sensitive insight into operational decisions, such as:

- **Demand forecasting and inventory planning:** Anticipating product movement by region or store, factoring in seasonality and local patterns
- **Dynamic pricing and promotion optimization:** Adjusting pricing based on demand signals, inventory position, and competitive conditions
- **Personnel planning:** Aligning staffing to expected traffic and workload patterns
- **Shrink reduction and loss prevention analytics:** Identifying where shrinkage occurs and focusing on interventions

Predictive systems work best when they can continuously collect data from stores and edge systems and sync insights across the enterprise. But this requires reliable data flow between the edge, the cloud, and IT HQ, without depending on a single connection or a fragile routing path.

Agentic AI

Autonomous systems in the store and supply chain

Agentic AI represents systems that can interpret objectives, plan tasks, and act, often coordinating across tools and environments. This shows up in retail environments as:

- **Autonomous replenishment:** "Intelligent shelves" or sensing systems that trigger restocks automatically when thresholds are met
- **Workflow agents for operations:** Tools that coordinate tasks such as ticket creation, diagnostics, dispatching, and follow-up across store networks
- **Robotics and automation:** In warehouses, where responsiveness and safety requirements are tight

Agentic systems require predictable latency, network resiliency, and — because they trigger action — strong segmentation. The network must support path diversity, so an outage doesn't stop automation, as well as policy controls that limit what automated systems can reach if something goes wrong.

Generative AI

Enhancing engagement and experience

Generative AI expands retail's ability to communicate with, guide, and deliver support for customers and employees. In practice, it is visible in areas such as:

- **In-app assistants and conversational commerce:** Customer Q&A, product guidance, and order support
- **Personalized recommendations:** Driving conversion and basket size through "did you forget...?" prompts and suggesting the right product add-ons
- **Associate enablement:** AI assistants that help new or seasonal staff perform more like experienced employees by surfacing answers to their questions
- **Content generation for marketing, planograms, and digital experiences:** Faster content updates with controls to keep content accurate, compliant, and on brand.

Generative AI often touches sensitive data such as customer profiles, loyalty signals, and camera imagery. That visibility raises governance requirements around access controls, segmentation, and consistent inspection, especially when a single weak endpoint could become an entry point for a bad actor.

Why AI-based applications are best executed at the edge

Retail creates "elephant flows" of data at the edge, and video is the largest source of such data. A single store can have dozens of cameras producing continuous streams. When AI is layered on top of video to detect suspicious behavior, track traffic patterns, or spot shelf gaps, for example, it becomes impractical to send all that data back to a cloud or data center for processing. The bandwidth costs are high, latency creeps in, and the insights arrive too late to be useful. That's why AI workloads like these need to run at the edge, close to where data is generated and actions occur.

In practice, edge AI in retail typically manifests in a few common architectural patterns. Predictive and computer vision AI are the most obvious fit for edge processing, especially for video analytics, loss prevention, safety monitoring, and operational awareness. Edge processing is often used for customer experience use cases that require instant responses, such as smart mirrors and virtual try-ons, where delays can ruin the experience. Agentic AI can appear at the edge as well when it orchestrates local actions, such as triggering an alert, opening a ticket, or adjusting a workflow based on what the system detects in real time.

Image courtesy of Getty Images

Most retailers don't choose between the edge and cloud; they adopt a hybrid approach. The edge handles immediacy and high-volume data reduction, turning raw video and sensor inputs into lightweight, actionable events. The cloud is used for the scale of centralized analytics, model training, and access to larger AI services. Central IT and HQ systems set the policy and governance standards and apply updates across hundreds or thousands of sites.

The key is that hybrid architecture only works when the network is built for it. Retail IT teams need upstream headroom to move insights and metadata without clogging the link. They also need consistent performance for real-time experiences and local failover that keeps the site running when connectivity degrades. Edge video analytics shows why. Local processing keeps raw streams on site, and the network carries only the insights upstream. Retail teams still rely on the WAN to deliver updates, enforce policy, and maintain centralized visibility. In other words, the edge makes AI workable at scale, and the network keeps it dependable.

5G and SD-WAN as the foundation for AI-driven retail

Retail IT leaders need a smarter, more resilient, and easier-to-operate network fabric that supports AI workloads at the edge without adding operational complexity.

5G

The catalyst for AI applications across sites, mobile, and IoT

In retail, 5G can connect sites and devices faster than wired service, and it works well in places beyond the main store, such as parking lots, pop-ups, warehouses, and delivery vehicles. Retailers can use 5G as their primary connection, a failover connection, or a temporary connection when they need to get online quickly.

Key advantages of 5G for retail AI include:

- Low latency and responsiveness for interactive experiences, real time analytics, and operational automation.
- Bandwidth designed for high-volume data streams such as video, guest Wi-Fi, and inventory management systems.
- Mobility and coverage, expanding the network "perimeter" to include vehicles and distributed assets.
- 5G standalone and slicing potential to prioritize essential traffic (e.g., transactions) during network congestion.

SD-WAN

The smart nervous system for AI-driven retail

As retailers scale, SD-WAN enforces application intent across every location, ensuring critical traffic behaves predictably even when conditions vary. In practice, that capability comes from four core functions:

- Application-aware routing identifies and prioritizes POS, voice, business-critical applications, and AI workloads based on performance requirements.
- Dynamic path selection continuously steers the most important and time-sensitive traffic onto the best available link, rather than relying on static failover.
- Single policy plane and visibility applies consistent segmentation and application policies across multiple sites and monitors performance end-to-end from one place, so teams can troubleshoot faster and enforce standards without bespoke, site-by-site configuration.
- Resiliency mechanics use failover, packet duplication, forward error correction, and WAN bonding to keep transaction and analytics workloads stable during link degradation.

All retail applications are not created equally: a guest Wi-Fi slowdown is inconvenient, but a POS interruption directly impacts revenue. With SD-WAN, IT teams can encode a hierarchy into policies that protect essential traffic and enforce them consistently across operations.

The combined value of 5G and SD-WAN for AI

While 5G expands the reach and agility of the WAN, SD-WAN provides the centralized control and policy enforcement needed for remote IT teams. Together, they:

- Reduce operational sprawl by managing more environments under a unified policy model.
- Extend consistent connectivity and security to stores and logistics, fleet, or warehouse environments.
- Maintain performance for important workloads while supporting new AI traffic patterns.

Best practices for integrating AI in retail

To integrate AI successfully, retail IT and networking teams should focus on three practical areas: containment through segmentation, scalable deployment through standardization, and policy-based controls that protect application performance.

Design for containment

As retail layers AI into their existing operations, segmentation becomes a requirement. POS systems, cameras, inventory systems, employee devices, and guest access should be isolated so that a threat or compromise in one zone doesn't affect another. This reduces risk and supports compliance expectations as more sensitive data is processed across the edge.

Standardize, then automate

Bespoke, site-by-site designs don't scale well in retail environments. Standardized templates, zero-touch provisioning, centralized policy, and automation make multi-site reliability realistic, especially under lean staffing constraints.

Plan for traffic competition

AI increases traffic variety with virtual chats, image and video, telemetry streams, and operational apps, all competing for priority. Plan for network pressure with application prioritization, routing policies, and continuous performance monitoring tied to user experience, not just bandwidth utilization.

Retail is moving toward more autonomous stores, agentic supply chains, and AI-assisted operations that run continuously across a distributed footprint. As these use cases expand, the ability to deploy, secure, and operate AI reliably at scale becomes a key requirement. This is best supported by repeatable network conditions, including standardized policy, clear traffic prioritization, and segmentation that reduces risk.

For retailers, being AI-ready starts long before AI is deployed. A modern network built on 5G and SD-WAN gives teams the operational confidence to support edge intelligence, protect sensitive data through segmentation, and scale innovation without compromising performance.

Learn more about enterprise wireless solutions

Takeaway: What you need to know about AI in retail

What's happening

Retail AI is moving from pilots into day-to-day operations across stores, warehouses, and fleets.

Why this matters

Once AI is tied to transactions, inventory accuracy, and frontline workflows, the network becomes the difference between repeatable results and uneven performance. AI puts more pressure on application experience, uptime, and security controls across many locations.

What you need

A network foundation that simplifies operations across sites and includes centralized visibility, consistent policy enforcement, and strong segmentation by system and risk level.

Why 5G and SD-WAN

5G adds deployment speed, mobility support, and WAN diversity (primary, backup, or rapid-deployment links). SD-WAN standardizes application-aware routing, segmentation, and end-to-end visibility across sites. Together they support a consistent operating model for AI across retail environments — without adding operational overhead.

Your takeaway

The near-term priority is to establish repeatable deployment conditions, including standard policies, clear traffic prioritization, and risk-containment segmentation. 5G and SD-WAN provide a practical foundation to scale AI in a controlled way as more sites and workflows come online.

Learn more about enterprise wireless solutions