Retail Practice

LLM to ROI: How to scale gen AI in retail

Retailers continue to experiment with generative AI and are seeing that the technology holds great promise for reigniting growth—if they move quickly to seize the opportunity.

This article is a collaborative effort by Alexander Sukharevsky, Andreas Ess, Denis Emelyantsev, Emily Reasor, and Holger Hürtgen, with Oleg Sokolov and Sergey Kondratyuk, representing views from QuantumBlack, AI by McKinsey, and McKinsey's Retail Practice.

August 2024

Once generative AI (gen AI) hit the mainstream, in late 2022, it took little time for retail executives to realize the potential in front of them. Mentions of artificial intelligence (AI) in retailers' earnings calls soared last year—which was no surprise, given that gen AI is poised to unlock between $240 billion to $390 billion in economic value for retailers, equivalent to a margin increase across the industry of 1.2 to 1.9 percentage points. This, combined with the value of nongenerative AI and analytics, could turn billions of dollars in value into trillions.

Over the past year, most retailers have started testing different gen AI use cases across the retail value chain. Even with all this experimentation, however, few companies have managed to realize the technology's full potential at scale. We surveyed more than 50 retail executives, and although most say they are piloting and scaling large language models (LLMs) and gen AI broadly, only two executives say they have successfully implemented gen AI across their organizations (see sidebar, "Our survey findings").

Some retailers have found it difficult to implement gen AI widely because it requires rewiring parts of the retail organization, such as technical capabilities and talent. Data quality and privacy concerns, insufficient resources and expertise, and implementation expenses have also challenged the speed at which retailers can scale their gen AI experiments.

Retail companies that have succeeded in harnessing gen AI's power typically excel in two key areas. First, they consider how gen AI use cases can help transform specific domains rather than spreading their resources too thin across a range of scenarios. Second, they effectively transition from pilot and proof-of-concept to deployment at scale. This requires not just data prioritization and technological integration but also significant organizational changes to support widespread AI adoption.

In this article, we explore which use cases can offer the most value and what organizational transformations are necessary to scale these technologies successfully.

Our survey findings

In April 2024, we conducted a survey of 52 global Fortune 500 retail executives. Our survey focused on the progress retailers made in exploring and experimenting with generative AI (gen AI). We found that most retail executives (90 percent) say they began experimenting with gen AI solutions and scaling priority use cases and that these experiments had knock-on effects across their other AI initiatives; two-thirds of retail leaders say they want to invest in and focus more on data and analytics.

Sixty-four percent of retail leaders say they have conducted gen AI pilots that have augmented their organizations' internal value chains, while 26 percent

say they are already scaling gen AI solutions in this area. What's more, 82 percent of retailers say they have conducted pilots for gen AI use cases related to the reinvention of customer service. Thirty-six percent say they are scaling gen AI solutions in this area.

Off-the-shelf gen AI tools have become more readily available in the past year. More than half of retail leaders surveyed (60 percent) opted for ready-made platforms, although the adoption rate of these third-party platforms is lower in areas such as procurement (18 percent) and commercial (25 percent). The adoption of third-party gen AI solutions will likely grow as the gen AI platform market

matures. Two-thirds of retailers say they intend to increase their gen AI budgets over the next year.

Meanwhile, 10 percent of retailers say they are adopting a wait-and-see approach to gen AI. They plan to integrate gen AI into their operations at a later date, particularly in the areas where there are no full-service platforms yet. This decision may stem from factors such as having insufficient expertise or organizational resources, issues related to data quality and privacy, as well as the expenses associated with gen AI implementation.

From the inside out: Two ways gen AI transforms retail

Retailers we spoke with have already piloted gen AI use cases within their internal value chains, and some are even beginning to scale gen AI solutions. Gen AI can help streamline operations, allowing leaders to make faster, better-informed decisions across retailers' internal value chains. The technology also offers immediate, no-regret efficiency gains, as well as applications that could redefine decision making in retail (more on this later).

Retailers have also experimented with gen AI to reinvent the customer experience. Gen AI can deepen relationships with customers (in part, by extending the interactions between retailers and customers across the customer journey) and help make the customer experience more personalized and fulfilling. The advanced conversational abilities of gen AI chatbots, powered by natural-language models, can make the smart-shopping assistant a primary shopping channel.

Augmenting retail's internal value chain
Gen AI has the potential to boost productivity and efficiency along each step of the retail value chain, including in marketing, commercialization, distribution, and back-office work (Exhibit 1).

Retailers can start to realize gen AI's impact across the value chain through quick-win use cases. These use cases generally require fewer resources to implement relative to their impact and compared with other gen AI use cases. In fact, retailers may more easily deploy current off-the-shelf tools without the need for much customization. Real examples of these use cases include the following:

— *Marketing.* Amazon launched an AI-powered image generation tool in late 2023 to help advertisers deliver a better ad experience. The tool uses gen AI text prompts to transform basic product photos into more realistic lifestyle images. For example, rather than showing a picture of a sofa against a white backdrop, AI can place the sofa in an AI-generated living room to help shoppers envision the product in a more relevant context. The tool so far has improved advertising click-through rates by up to 40 percent.

— *Software development.* "Copilots," or gen AI tools that help employees do their jobs by providing a starting point for a task, can boost tech talent productivity by reducing the time spent on software engineering tasks by up to 60 percent. Mercado Libre deployed some of these copilots to improve satisfaction and productivity among the company's development teams, empowering them to focus on higher-value work by automating more repetitive tasks.

Gen AI has the potential to boost productivity and efficiency along each step of the retail value chain.

Exhibit 1

There are several touchpoints where generative AI can transform the retail value chain.

Retail value chain

	Before generative AI	After generative AI
Procurement	All supplier negotiations (including end-to-end contract creation) handled manually by associates, often leading to overlooked details Tedious supplier assessments based on limited data, leading to suboptimal choices	The initial round of supplier negotiations handled by generative AI (gen AI) chatbots Procurement associates utilize gen AI to assist in closing deals (eg, gen-AI-powered briefs and automated summaries of supplier terms and key insights)
Distribution	End-to-end communication with third-party logistics handled by associates Delayed response to distribution disruptions due to complexity of supply chain operations	Initial communication and email messages to third-party logistics handled by gen AI chatbots Returns management process, along with a response to distribution disruption, supported by gen AI
In-store operations	Information searches (eg, price, in-store location, stock level) handled manually by associates, leading to delayed customer service	Associates use gen-AI-powered assistants for instant voice access to information (eg, prices and promos, product location, stock level)
E-commerce	Hundreds of hours spent on the generation of e-commerce content Manual rule-based website personalization, consuming employees' resources	Automated generation of e-commerce content (eg, product profiles, descriptions) within a few minutes E-commerce customer experience personalized spontaneously by automated front-end development techniques
Marketing	One-size-fits-all marketing approach due to limited customer insights derived from structured data Creation of marketing materials through a lengthy, iterative process	Unlimited insights extracted from different unstructured sources (eg, product reviews) Fully personalized marketing materials generated with increased efficiency for every customer
Back office	Time-consuming administrative processes, such as HR and payroll, prone to errors and inefficiencies	The next-generation "white collar" lean—transferring administrative processes of support functions to gen-AI-powered chatbots and interfaces (eg, software development copilots, HR/financial copilots)
Commercial	Analytical tools of different maturity level, sometimes hard to adopt	Enhancement of analytical tools with gen AI interface, automation of creative tasks (merchandiser copilot)
End-to-end value chain	Independent decision making by individual functions, leading to a recurring cycle of searching for the underlying causes of commercial events, often failing to identify the true underlying factors	End-to-end optimization of decision making across the value chain through faster, more precise, and personalized insights from structured and unstructured data sources, fueling decision on pricing, promotions, stock allocation, digital marketing, and other levers of performance

McKinsey & Company

— *In-store operations.* In June 2023, Lindex, a Swedish retailer, announced the release of the "Lindex Copilot" to support its store employees. The tool, which is trained on the company's sales and store data, provides employees with personalized advice and guidance about store operations and information about daily tasks.

While the above examples can help simplify daily tasks, gen AI can also help retailers accelerate their decision making by automatically generating insights, root causes, and domain-level and company-wide responses (Exhibit 2).

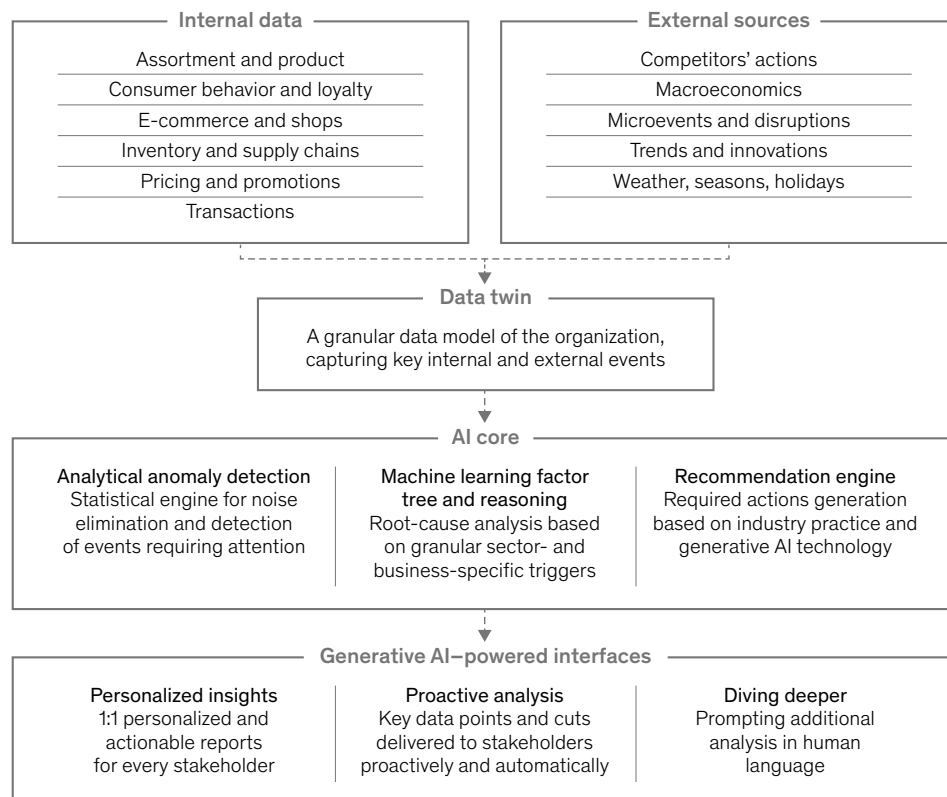Retail operations are affected by countless forces that are difficult to quantify and track,

making performance analytics and forecasting an arduous task. Traditionally, teams might spend weeks studying competitors' tactics, changes in pricing and promotion, supply chain issues, and unexpected disruptions to understand sales declines and devise strategies to avoid future sales drops. The combination of gen AI and advanced analytics can revolutionize this process: rather than manually assessing that data, workers from across the company—from CEO to category manager—can access a personalized report featuring key performance insights and suggested actions.

Let's use a hypothetical electronics retailer as an example. The retailer's television sales are 6 percent lower than it had forecasted. The retailer's team

Exhibit 2

Generative AI can help bring clarity to retail decision making.

Internal and external inputs needed to enable generative AI decision making

Internal data	External sources
Assortment and product	Competitors' actions
Consumer behavior and loyalty	Macroeconomics
E-commerce and shops	Microevents and disruptions
Inventory and supply chains	Trends and innovations
Pricing and promotions	Weather, seasons, holidays
Transactions	

Data twin
A granular data model of the organization, capturing key internal and external events

AI core

Analytical anomaly detection	Machine learning factor tree and reasoning	Recommendation engine
Statistical engine for noise elimination and detection of events requiring attention	Root-cause analysis based on granular sector- and business-specific triggers	Required actions generation based on industry practice and generative AI technology

Generative AI–powered interfaces

Personalized insights	Proactive analysis	Diving deeper
1:1 personalized and actionable reports for every stakeholder	Key data points and cuts delivered to stakeholders proactively and automatically	Prompting additional analysis in human language

McKinsey & Company

spent a week looking for the root cause of the decline and came up with a dozen potential reasons: Could the missed sales forecast have been caused by the unusually rainy weather? A delayed product release? Or were temporary out-of-stock items and a weak promotional campaign to blame?

In this example, a gen AI system, trained on the retailer's proprietary data, could automatically analyze the impact of not only these potential root causes but also additional scenarios, such as what actions its competitors may have taken at the same time. A cross-functional team, led by the retailer's technology leaders and considering input from sales and commercial teams, could work with technology providers to customize the retailer's AI- and gen-AI-powered system. The gen AI platform could then create a list of causes by impact, as well as a set of actions the retailer could consider to help reduce sales drops in the future.

Based on our early work with retailers, we expect gen-AI-powered decision-making systems to propel up to 5 percent of incremental sales and improve EBIT margins by 0.2 to 0.4 percentage points.

When it comes to using gen AI copilots, companies will need to decide if they are a "taker" (a user of preexisting tools), a "shaper" (an integrator of available models with proprietary data for more customized results), or a "maker" (a builder of foundation models). Across the internal value chain, most retailers will likely adopt the taker archetype, using publicly available interfaces or APIs with little to no customization to meet their needs.

However, many of today's off-the-shelf solutions don't offer the functionality that some retailers need to fully realize the technology's value, since the technology powering these solutions typically doesn't account for sector- and company-specific data. At the same time, most retailers won't be able to adopt the maker archetype, given that the costs associated with building foundation models are outside the typical retailer's budget. In these cases, retailers may opt for the shaper archetype, customizing existing LLM tools with their own code and data. The shaper archetype will also be relevant for gen AI decision-making use cases.

How many resources a retailer invests in shaping its gen AI tools will depend on the market it intends to serve, which use cases it wants to prioritize, and how these use cases complement the retailer's core value proposition.

Reinventing the customer experience
Today, retailers typically engage in only three of the seven steps of the customer journey. Gen AI has the potential to increase retailer engagement and reinvent the customer experience across the entire customer journey (Exhibit 3).

Gen-AI-powered chatbot assistants are one primary tool retailers can use to better engage with customers. Customers can use chatbots to receive product recommendations, learn more about a product or retailer, or add or remove items from their virtual shopping carts. Importantly, since many consumers will use these chatbots before deciding to purchase a product rather than after, using chatbots allows retailers to engage with customers earlier in their shopping journey, which can help increase customers' overall satisfaction.

Gen AI chatbots work by recognizing the intent of a customer's message. An LLM agent—the system that the chatbot relies on for its reasoning engine—processes the customer's message and is then connected to various data sets (such as a retailer's SKU base) and to other models, such as an analytical personalization engine. To create the best outputs, a retailer must dedicate resources toward product design and conduct frequent user testing to calibrate how it wants the chatbot to process the customer's message. (How customers most frequently use the chatbot will largely determine this calibration.)

For example, a shopper might be interested in planning a dinner party but may not know what to buy. After the customer provides the gen AI assistant with a few details about the dinner party— such as how many people are attending, whether any guests have dietary restrictions, and overall budget—the gen AI assistant could provide specific product recommendations based on the customer's preferences or purchase history.

Exhibit 3

Generative AI can make the customer journey more efficient—for both retailer and shopper.

Retail customer journey

	Before generative AI	After generative AI
Purchase idea and context	Manually searching across multiple sources, or ideating independently	Customer provides brief context Generative AI agent understands the intent and offers suggestions based on preferences
Creation of purchase list	Manually creating the purchase list, relying on customer memory or experience	Generative AI agent provides an accurate shopping list that includes the items/ingredients needed for the purchase idea
Selection of preferred retailer	Choosing based on past experiences or spontaneity—hard to control	Generative AI minimizes switching, as the customer is already shopping in the retailer's app
Search for required item	Time-consuming browsing through multiple categories and filters online to find the right SKU	Generative AI intelligently navigates through catalogs to find items that perfectly match the customer's purchase history and preferences
Selection of item and quantity	Manually comparing each product across multiple dimensions (eg, price, quality, reviews and ratings, size)	Adaptation of basket for customer needs, including smart comparisons of products, considering factors from unstructured data (eg, style, customer and market reviews)
Order confirmation, delivery date selection	Navigating through multiple checkout pages, entering delivery and payment information, and choosing from limited delivery slots that may not align with the customer's schedule	Confirmation of order and specifying delivery preferences in a conversational style (eg, "Please deliver this order to my mother's address on any day except Wednesday")
Product usage	Analyzing online reviews and web forums to gain insights into product usage methods	Recommendations for videos, production usage insights, and step-by-step instructions, tailored to customer context

McKinsey & Company

While chatbots can be a convenient tool to help reduce customers' mental load and shopping time, to truly transform the shopping experience and win over customers, chatbots will need to be deeply personalized—for example, being able to remember customers' order histories, product preferences, and shopping habits. Many leading retailers, particularly in the grocery and fashion spaces, have already begun experimenting with chatbots, though most of these early experiments have not yet harnessed the power of personalization (see the sidebar "Retailers embark on the chatbot journey").

As is the case with internal value chain gen AI use cases, retailers often adopt the "shaper" archetype for gen AI use cases that transform the customer experience.

Determining the costs of chatbots in retail. The first concern many retailers have about integrating gen-AI-powered chatbots into their business is how much it will cost. That depends on a few factors. Product performance metrics (or the length of a conversation between a customer and chatbot) is one of the first considerations. The length of the conversation is inversely related to the quality of personalization—meaning, the more personalized a chatbot is for a given customer, the shorter their conversation. Purchase conversions are another factor. The higher the conversion rate, which is linked to the effectiveness of the chatbot, the lower

Retailers embark on the chatbot journey

While some retailers have adopted a wait-and-see approach to generative AI (gen AI), others have already started experimenting with the technology (exhibit). Leading retailers, particularly those in grocery and fashion, started experimenting with gen-AI-powered chatbots in late 2023. These chatbots have taken different forms. Walmart launched its "Text to Shop" tool, where customers can text the retailer to search for items, add or remove items from their carts, reorder products, and schedule deliveries. Instacart created a ChatGPT plug-in that allows users to plan a meal in ChatGPT and then convert the output into a basket on Instacart's website.

Exhibit

Retailers, particularly those in grocery and fashion, have started experimenting with generative AI–powered conversational chatbots.

Retailer implementation of AI-powered chatbots, nonexhaustive

	Implementation	Retailer	Retailer category	AI-powered chatbot type
1	Dec 2022	Walmart	Grocery	Messenger[1]
2	Mar 2023	Instacart	Grocery	ChatGPT plug-in
3	Apr 2023	KNXT	Fashion	Website/app
4	Apr 2023	Zalando	Fashion	Website/app
5	June 2023	Carrefour	Grocery	Website/app
6	Nov 2023	Castorama	DIY	Website/app
7	Jan 2024	Walmart[2]	Grocery	Website/app
8	Feb 2024	Amazon	Grocery	Website/app
9	Feb 2024	IKEA	DIY	ChatGPT plug-in
10	Feb 2024	Ssense	Fashion	ChatGPT plug-in
11	Mar 2024	Lyko	Beauty	Website/app
12	Reported plans	Victoria's Secret	Fashion	Website/app

[1]Eg, iMessage, Instagram direct message, WhatsApp.
[2]SmartSearch functionality in the app; chatbot still available only in iMessage.

McKinsey & Company

the net operational costs of that chatbot. A third factor is the price of LLM APIs. The cost of using these LLM APIs has dropped dramatically in the past year (for example, when comparing the cost of input tokens, GPT-4o, released in May 2024, is half as expensive to operate as GPT-4 Turbo, released a year earlier). AI experts believe that the price of LLM APIs will continue to drop substantially, with some estimates showing a drop of as much as 80 percent within the next two to three years.

Based on our experience building gen AI chatbots with retail companies across a range of realistic scenarios, a 2 to 4 percent basket uplift can justify LLM costs. Retailers can also combine the power of their generative and analytical AI products to

further justify LLM costs. For example, companies can first use gen AI to learn more about a customer, then use analytical models to surface personal offers relevant to that customer. Together, these two technologies can help increase sales conversions.

When building a business case, retailers should also consider the investment required to develop a chatbot. Sometimes, the basket uplift may not be high enough to cover the cost of the investment. To understand the full return on their investment, retailers should factor in the cost of attracting new customers who will use these tools, as well as how much the tool can increase the purchase frequency for existing customers.

Measuring chatbots' impact. In controlled customer experiments, we've seen chatbots create a significant increase in convenience for customers. When comparing a traditional retailer app with the minimum viable product of a gen-AI-enabled chatbot, the chatbot reduced the time spent to complete an order by 50 to 70 percent.

Retailers that aren't ready to invest in chatbots may instead choose to launch smart-search functionality. Smart-search tools allow a customer to receive a list of recommended products by asking a question rather than needing to engage in a conversation with a chatbot. (For example, a customer might search for "dinner party supplies," and the smart-search tool would provide a list of products that one might need for a dinner party). While traditional search uses basic algorithms and relies on keyword matching, smart-search tools powered by gen AI can better understand the context and intent of a search term, even if it veers away from keyword use. Although these smart-search tools may be limited in functionality compared with using chatbots—and therefore limited in impact—they are easier and less expensive to develop. They also carry fewer risks compared with using a chatbot; their outputs are generally limited to a list of products rather than longer text that a chatbot would give, which means the responses are less likely to be harmful, offensive, or inaccurate.

How to scale the use of gen AI in retail

Gen AI is no longer a novelty. As companies figure out how to implement the technology to create real value, best-in-class retailers will need to move from testing to scaling or else risk falling behind their competitors—or, worse, losing customers. To scale their gen AI tools, retail executives can consider five imperatives for outcompeting in digital and AI:

— *Identify domain-level transformation candidates.* Retail executives should identify the various domains where a transformation is needed, such as in customer experience, marketing, or store employee productivity, before they identify which gen AI use cases to pursue. By identifying transformations at the domain level first, retailers can determine which tools will bolster gen AI's impact, such as robotic process automation (RPA) or advanced analytics.

— *Upskill talent to develop gen AI skills.* Retail people leaders should offer both technical and nontechnical talent the opportunity to engage in learning programs, such as those focused on gen AI software development and prompt engineering.

— *Form a centralized, cross-functional team to enable scaling.* While scaling gen AI will hinge on a retailer's tech capabilities, retailers can gather leaders from across the organization to identify how gen AI can help improve the business. These cross-functional teams, convened to help accelerate scaling in the short term, should have shared goals that extend across the business and reinforce the retailer's overall strategy.

— *Set up technology architecture to scale.* Before committing to a specific gen AI vendor, retailers should experiment with different ones to assess which vendor will best suit their needs. The ideal gen AI architecture for retailers will be agile enough to make switching between LLMs easier to do, thus making scaling the technology across the organization easier as well. (This means using modular components that can be easily swapped out.)

— *Ensure data quality to fuel models.*
Unstructured data will be critical to powering retailers' gen AI tools and providing key customer insights. Retailers should identify the unstructured data sources that differentiate them from other retailers (grocers, for example, could develop new recipe databases or leverage existing ones) and establish metadata tagging standards so tech teams can more efficiently power a retailer's gen AI models. Decisions on data should be backed by a clear understanding of the data's business application.

Some of the guidance outlined above may be sector-agnostic, but scaling gen AI in retail is unique because several of the technology's use cases involve direct interactions with consumers. In retail, even a 1 percent margin of error could result in millions of customer-facing mistakes. This emphasizes the importance of strong gen AI risk guidelines and safety testing. The stakes may be higher, but the rewards are, too.