



# DEMYSTIFYING AI FOR KEY PLAYERS IN CPG: BRANDS, RETAILERS, DISTRIBUTORS, AND BROKERS

WHITE PAPER

**MAY 2024**

# Table of Contents

Executive Summary -----	3
Introduction to Artificial Intelligence (AI) and Machine Learning (ML) in CPG and Retail ---	4
Evolution to Generative AI -----	5
Cautionary Tale of Gen AI -----	6
Comparative Analysis of Large Versus Small Language Models in CPG and Retail -----	8
What is an AI Hallucination and How to Avoid Them? -----	10
First-Party Data Examples in CPG and Retail -----	11
Leveraging Third-Party Data in CPG and Retail -----	13
Steps to Bringing it All Together -----	14
Conclusion -----	15
Authors -----	16

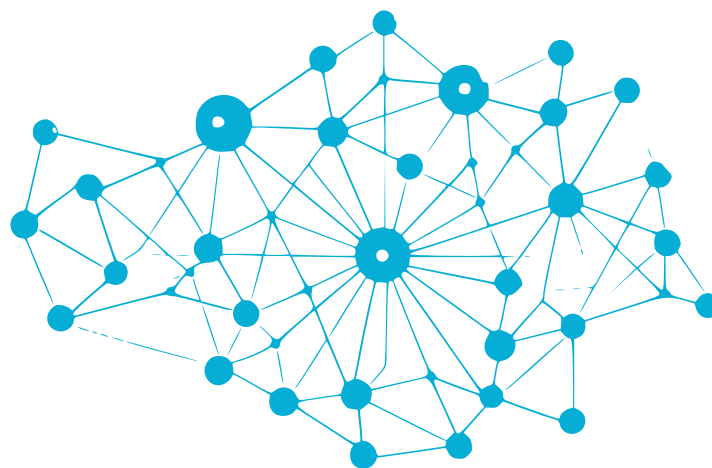


# Executive Summary

In today's rapidly evolving Consumer Packaged Goods (CPG) industry, stakeholders are increasingly turning to Artificial Intelligence (AI) for a competitive edge. The application of AI technologies – including Machine Learning (ML), Predictive Analytics, and Natural Language Search (NLS) processing – is revolutionizing traditional practices, offering unprecedented efficiency, and insights.

This whitepaper offers a point of view based on the authors' collective experience in the CPG ecosystem and the dynamic world of AI, ML, Generative AI, Large Language Models (LLMs), Small Language Models (SLMs), AI hallucinations, and first-party, and third-party data. The goal is to demystify these key concepts, offering clarity on their intricacies and revealing how Generative AI is reshaping the CPG landscape.

By bringing together data analytics and predictive intelligence through solutions like the Aisle AI Data Intelligence Platform, CPG and retail leaders can expedite new product launches, understand brand performance, and extract valuable insights to enhance profitable retail strategies.



Confidential © Aisle AI All rights reserved.

# Introduction to Artificial Intelligence (AI) and Machine Learning (ML) in CPG and Retail

Artificial Intelligence (AI) integration is revolutionizing the CPG industry, impacting customer interactions, operations, and strategic decisions. With AI and Machine Learning (ML) at its core, CPG businesses can leverage advanced algorithms and techniques to analyze vast datasets, predict consumer behavior, and optimize supply chains. ML, a subset of AI, focuses on refining algorithms for autonomous decision-making through a process called training. In ML, algorithms are trained on large datasets to learn patterns, trends, and relationships within the data.

Conventional AI and ML methods are widely used in demand forecasting, customer segmentation, pricing optimization, and supply chain management. Industry leaders apply these techniques to analyze data and enhance processes. For instance, sales teams utilize ML for demand forecasting, while supply chain teams employ predictive analytics for inventory management.

AI finds multifaceted applications in the CPG landscape:

- **Marketing and Sales:** AI-powered analytics dissect consumer data from various channels, aiding in the creation of personalized marketing campaigns and promotions to boost engagement and sales.
- **Merchandising and Assortment Planning:** AI algorithms analyze sales data and market trends to optimize product assortments, placements, and pricing strategies, ensuring maximum sales.
- **Supply Chain Management:** AI optimizes inventory management, demand forecasting, and logistics operations, minimizing costs and disruptions while maintaining optimal inventory levels.
- **Customer Experience Enhancement:** AI-driven chatbots and virtual assistants deliver personalized customer support and recommendations, while sentiment analysis improves product offerings and tailors experiences.
- **Loss Prevention and Fraud Detection:** AI surveillance systems monitor store activities, identifying anomalies indicative of fraudulent activities or theft, bolstering security, and reducing losses.
- **Predictive Analytics and Forecasting:** AI and ML models forecast sales, trends, and market segments, enabling retailers to anticipate demand and make informed decisions.

AI and ML provides CPG brands with deeper insights into consumer preferences, facilitating targeted marketing, optimized product assortments, and personalized promotions ultimately boosting efficiency, profitability, and competitiveness. Retailers and distributors benefit from AI-driven inventory management, demand forecasting, and pricing strategies, leading to improved profitability and customer satisfaction.

# Evolution to Generative AI

The latest iteration of AI to hit the market is Generative AI or “Gen AI”, a subset of AI that focuses on generating new data “predictions” based on the information it learns over time. Unlike traditional ML algorithms that are primarily focused on making predictions or classifications based on existing data, Gen AI creates new data samples that exhibit similar characteristics to the training data. The most well-known Gen AI applications OpenAI’s ChatGPT and Google’s Gemini, are based on building and training Large Language Models (LLM).

Gen AI distinguishes itself from traditional AI and ML by empowering the creation of new data samples, presenting opportunities for CPG companies to enhance customer engagement and operational efficiency. Moreover, Gen AI offers unique advantages to the CPG industry, allowing leaders to assess labor versus shrinkage across stores, propose inventory adjustments, reallocate resources, and bolster security measures in new store areas.

Additionally, Gen AI has the capability to generate images for product catalogs and tailor marketing campaign descriptions to specific demographics.

In comparison to both traditional AI and ML, Gen AI has some distinct characteristics and applications:

**Creation of New Data:** Traditional AI and ML focus on analyzing existing data for tasks like classification and regression, while Gen AI excels at generating entirely new data samples. This capability is valuable in creative applications such as image and text synthesis, music composition, and art generation. For instance, Gen AI can create realistic images, compose music, and mimic human writing.

**Unsupervised Learning:** Gen AI often involves unsupervised learning techniques, where the model learns to represent the underlying structure of the data without explicit labels or supervision. This allows Gen AI models to capture complex patterns and dependencies in the data and generate new samples that are consistent with the learned distribution.

**Data Augmentation:** Gen AI can also be used for data augmentation, where new synthetic data samples are generated to augment the training dataset. This helps improve the robustness and generalization of machine learning models, especially in scenarios where the training data is limited or imbalanced.

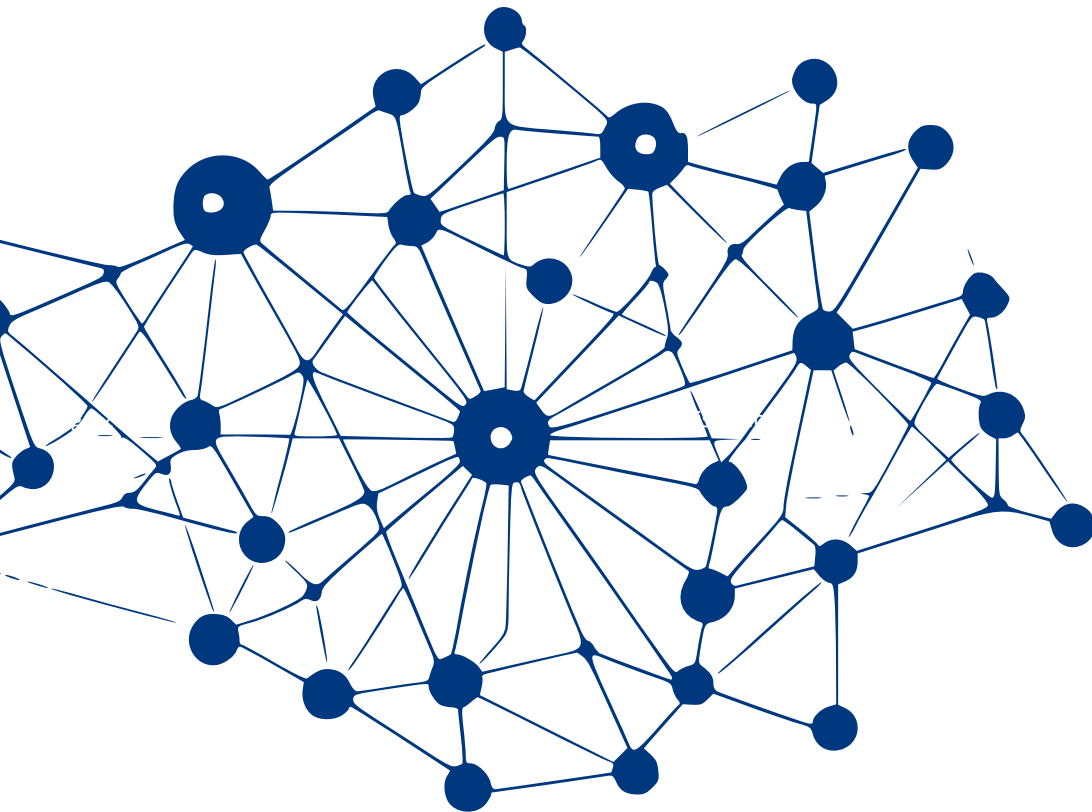
**Adversarial Training:** Gen AI often involves adversarial training, where two neural networks, known as the generator and the discriminator, are trained simultaneously in a competitive manner. The generator aims to generate realistic data samples, while the discriminator aims to distinguish between real and generated samples.

# Cautionary Tale of Gen AI

Harnessing the power of Gen AI models can pose significant hurdles unless you have access to specialized data scientists proficient in SQL, Snowflake, Python, and ETL (Extract, Transform, Load data integration process) transformation tools. Unlike traditional programming, ML algorithms evolve through data patterns, continuously improving performance and adapting to complex scenarios, making them essential assets in the dynamic CPG landscape. Here are a few key challenges that may affect how you deploy Gen AI to benefit your CPG business:

- **Accessing Generative AI Models:** This entails either developing in-house proficiency in AI development or collaborating with AI technology providers. Depending on the scale and complexity of AI requirements, this may necessitate recruiting AI specialists, data scientists, or enlisting the services of external AI providers.
- **Data Scientists and Analysts:** These experts can play a vital role in comprehending stored data, pinpointing pertinent datasets for AI model training, and evaluating outputs produced by the AI. Their responsibility is to ensure effective AI training on the data and extract valuable insights from the generated content.
- **Data Warehouse:** Having access to the data warehouse platform provided by Snowflake or Databricks is crucial for storing and managing substantial amounts of data. This necessitates proper licensing and infrastructure configuration to guarantee secure data storage and efficient accessibility for AI applications.
- **Integration and Development Teams:** Enterprises require proficient IT experts capable of integrating generative AI models with data warehouses. This could entail crafting custom APIs, connectors, or workflows to extract data, input it into the AI models, and store the resulting outputs.
- **Compliance and Legal Expertise:** With the sensitive nature of data stored in platforms like Snowflake and Databricks, coupled with the potential ramifications of using AI-generated content, companies require legal and compliance professionals. They ensure that data usage aligns with pertinent regulations (such as GDPR or CCPA) and ethical standards.
- **Training and Education:** Employees require instruction on effectively utilizing generative AI models and interpreting their outputs. This could entail workshops, training sessions, or educational materials to familiarize users with the capabilities and constraints of AI-generated content.
- **Quality Assurance and Monitoring:** Setting up procedures for quality assurance and monitoring is vital to guarantee that AI-generated content meets the desired benchmarks for accuracy, relevance, compliance, and the avoidance of AI hallucinations. This might entail deploying automated checks, human review processes, and feedback mechanisms to iteratively enhance AI outputs.

As evident, it's quite a significant endeavor. However, by effectively leveraging these business resources, organizations can tap into the potential of Gen AI models and data to foster innovation, gain valuable insights, and enhance decision-making processes.



Confidential © Aisle<sup>AI</sup> All rights reserved.

# Comparative Analysis of Large Versus Small Language Models in CPG and Retail

While most Gen AI applications today are based on Large Language Models (LLM), it is also possible to train Small Language Model (SLM) AI applications focused on more a narrowly defined scope with curated content. The distinction between a LLM trained across numerous industries and a SLM specialized in CPG and retail data points is significant, with differences manifesting mainly in the scope, focus, and depth of understanding relevant to their respective domains.

## **Large Language Model (LLM) Across Every Industry**

Large Language Models (LLMs) like GPT-3 are distinguished by their exposure to a diverse range of data from fields such as finance, healthcare, technology, and law. This broad data exposure equips them with a comprehensive understanding of general concepts, knowledge, and linguistic nuances across various industries. However, while they cover a wide breadth of topics, LLMs tend to offer a generalist view that may lack depth in specific areas, including the CPG and retail sectors. Their extensive training allows for a wide-ranging yet not deeply specialized knowledge of any singular field.

The generalist nature of LLMs enables them to produce content that spans across different sectors, yet the specificity and nuanced insights required for the unique challenges and demands of the CPG and retail industry might not always be met with precision. An intrinsic feature of LLMs is their continual learning from data and interactions inputted during usage, which can raise concerns for businesses wary of unwittingly sharing proprietary information or innovative ideas into the public domain. This ongoing incorporation of new data means LLMs are perpetually learning and updating, potentially blurring the lines between private innovation and public knowledge.

Confidential © Aisle<sup>AI</sup> All rights reserved.



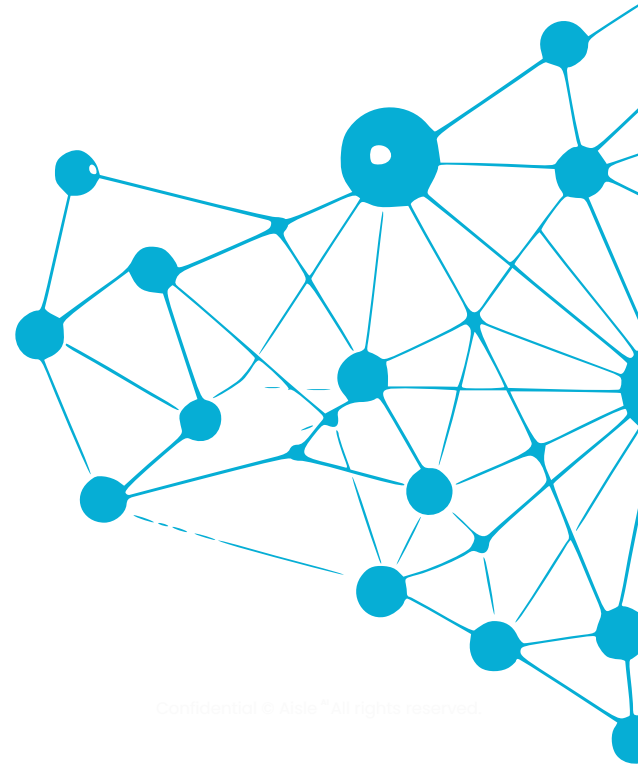
### **Small Language Model (SLM) Trained on CPG and Retail Data**

Alternatively, Small Language Models (SLMs) are meticulously trained using data relevant to the CPG and retail sectors. This specialized training results in a model finely attuned to the specific terminology, nuances, and unique challenges prevalent within these industries. While SLMs have a narrower scope across industries, if properly trained their ability to understand CPG and retail intricacies is unparalleled.

Designed to interpret and process CPG and retail-centric data, SLMs possess an in-depth understanding of industry dynamics, consumer behavior, supply chain complexities, and marketing strategies. This specialized dataset empowers SLMs to deliver responses that are not only relevant but also contextually precise, addressing the nuanced demands of the CPG and retail sectors.

Moreover, SLMs can operate in private environments, ensuring that proprietary content, data, innovations, and ideas remain confidential and inaccessible to competitors. Unlike LLMs that are trained by a wide range of users and may return varied responses over time, SLMs offer consistent and accurate results. By being specifically trained and fine-tuned for the CPG industry, SLMs minimize the risk of generating incorrect responses or hallucinations, thereby instilling confidence in the analytics and predictions they provide.

The decision to utilize a large, industry-agnostic language model or a small, domain-focused model depends on the task's nature. While LLMs are suitable for generalist purposes due to their broad knowledge base, the precision and depth of understanding offered by SLMs tailored to CPG and retail provide trustworthy, invaluable insights for stakeholders within these industries.



Confidential © Aisle AI All rights reserved.

# What is an AI Hallucination and How to Avoid Them?

AI models extract insights from training data, where biases or inaccuracies can cause skewed outputs, referred to as hallucinations, especially in LLMs. Hallucinations occur when a Generative AI model produces inaccurate information, masquerading it as true, stemming from limitations or biases in the model's training data and algorithms.

A lack of diversity in training data may hinder AI's interpretation of unfamiliar inputs, leading to hallucinations. Complex real-world scenarios occasionally challenge AI interpretation, resulting in hallucinations when faced with ambiguous data. Preventing AI hallucinations requires a comprehensive approach encompassing meticulous data management, model training, and user engagement.

Here are some effective strategies to address these concerns:

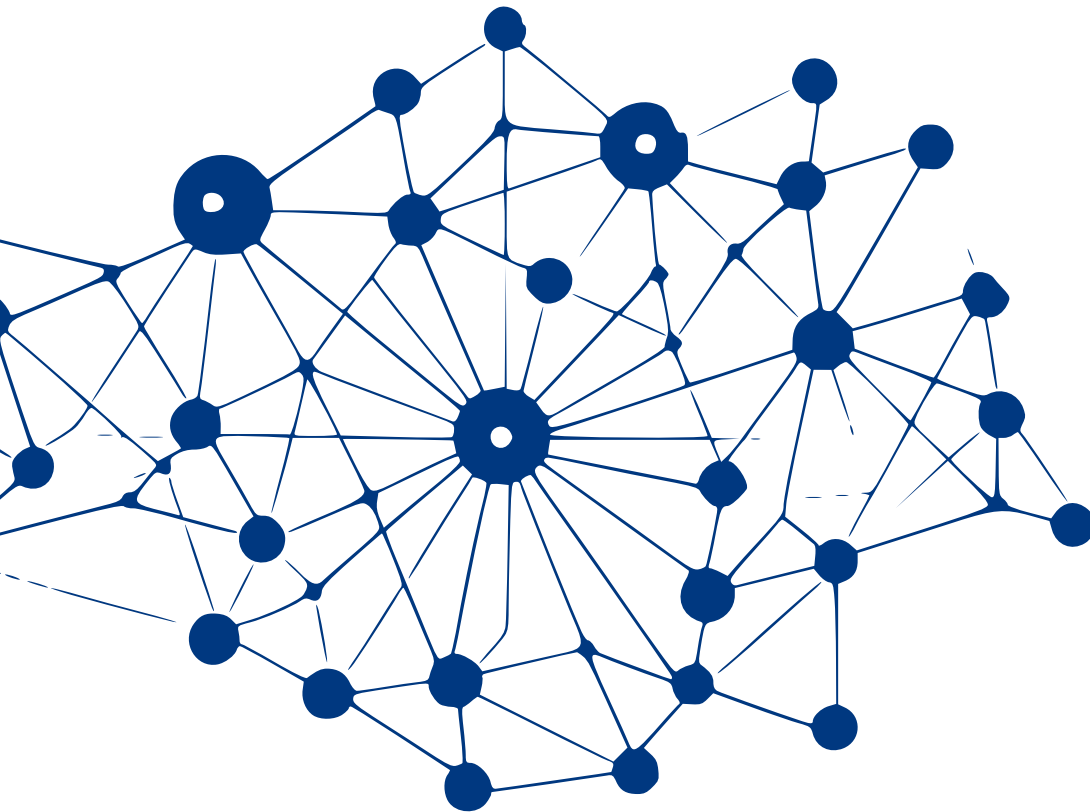
- **Use High-Quality Data:** Ensure the AI's training relies on top-notch, diverse, and representative datasets to minimize bias and inaccuracies.
- **Regular Updates:** Keep the AI's knowledge fresh with periodic updates to prevent it from relying on outdated information, enhancing its adaptability and relevance.
- **Implement Safety Measures:** Set up strict guidelines and safety mechanisms to prevent the AI from generating potentially misleading or harmful content, thereby boosting its reliability and safety.
- **Provide Clear Prompts:** Develop clear and concise prompts to guide the AI towards producing accurate and relevant outputs, aligning with user expectations and intentions.
- **Incorporate “Humans in the Loop”:** Introduce human oversight to review the AI's outputs, particularly in sensitive domains like healthcare or law enforcement, ensuring accountability and ethical integrity.
- **Conduct Adversarial Testing:** Test the AI rigorously with diverse and challenging scenarios to strengthen its resilience against unexpected inputs and reduce the likelihood of hallucinatory responses.
- **Utilize Grounding Techniques:** Employ grounding techniques to connect the AI's responses to verifiable information sources, enhancing its precision and reliability.

# First-Party Data Examples in CPG and Retail

Let's switch gears and talk about first-party data in CPG and retail. This critical data refers to information collected directly from customers or users or created by the companies themselves. It is data that the organization owns and has gathered through interactions with customers across various touchpoints, such as purchase history, customer feedback, and loyalty program data, offering deep insights into consumer behavior and preferences. Here are some examples of first-party data in the CPG and retail industry:

- **Purchase History:** Information about customers' past purchases, including product categories, specific items bought, purchase frequency, and transaction amounts. This data is typically collected through point-of-sale (POS) systems in physical stores or e-commerce platforms for online purchases.
- **Customer Profiles:** Profiles created for individual customers or segments based on demographic information, such as age, gender, location, income level, and household size. These profiles may also include psychographic data, such as lifestyle preferences, interests, and purchase motivations.
- **Website and App Interactions:** Data collected from customers' interactions with the company's website, mobile apps, or other digital channels. This includes information such as pages visited, products viewed, time spent on site, search queries, and actions taken (e.g., adding items to cart, completing a purchase).
- **Loyalty Program Data:** Information gathered from loyalty programs, including member profiles, points earned and redeemed, rewards preferences, and participation in promotional activities. Loyalty program data provides insights into customer behavior and preferences, as well as opportunities for targeted marketing and personalized offers.
- **Email and Marketing Campaign Data:** Data collected from email marketing campaigns, including open rates, click-through rates, conversion rates, and engagement metrics. This information helps measure the effectiveness of marketing efforts and tailor future communications to customer preferences.
- **ERP and Operational Data:** Data that is created by transaction systems in the normal course of business such as shipment, depletions, and inventory data.
- **Sales and Marketing Information:** Information created in the sales and marketing processes such as product detail, pricing detail, promotions calendar, and trade spending.
- **Broker / Distributor Data:** Data provided by brokers and distributors back to CPG manufacturers reflecting what is happening in the retail markets including sell-in and sell-out for specific products and outlets.

First-party data is valuable for CPG companies because it provides direct insights into their customers' behaviors, preferences, and needs. By leveraging first-party data effectively, companies can personalize marketing campaigns, improve product offerings, optimize pricing strategies, enhance customer experiences, and drive business growth. Additionally, first-party data is often considered more reliable and trustworthy than third-party data, as it comes directly from the source and reflects the company's own interactions with its customers.



Confidential © Aisle<sup>AI</sup> All rights reserved.

# Leveraging Third-Party Data in CPG and Retail

Another important dataset is known as third-party data. Third-party data comes from external sources not directly connected to the company and offers a diverse range of valuable information that complements our internal data sets. Incorporating third-party curated and syndicated data can significantly enrich the breadth and depth of insights we can generate for the CPG and retail industry. Examples of third-party data integral to operations include:

- **Syndicated Data:** Purchase and category data purchased from major consumer syndicated data providers such as NIQ, Circana, and SPINS. This data is commonly used, most often in a stand alone environment or dashboard, to inform sales and marketing activities.
- **Demographic and Geospatial Data:** From the U.S. Census Bureau and Experian, this data aids in understanding the demographic and geographic contours of target markets, which is vital for tailoring marketing efforts and optimizing store locations.
- **Credit Card Transaction Data:** Analyses of purchase patterns and preferences, providing a real-time pulse on consumer spending that aids in demand forecasting.

- **Weather and Environmental Data:** Information from The Weather Company and NOAA assists in adjusting marketing and inventory strategies to align with seasonal consumer behavior and preferences.
- **Marketplace Data:** Insights from Amazon and eBay about consumer reviews and competitive dynamics offer a comprehensive view of the e-commerce landscape, contributing to more effective online strategy formulations.
- **Market Research Reports:** Comprehensive analyses from leading research firms that provide insights into market trends, competitive landscapes, and consumer behavior, thereby enabling more informed strategic decisions.

Utilizing third-party data through solutions like what Aisle AI offers provides a more complete understanding of the market, empowering CPG businesses to not only react to current trends but also to anticipate future shifts. The key to being able to mix these different data sets to derive powerful insights while simultaneously maintaining strict security protocols and IP is to use a "clean room". In the clean room environments that Aisle AI uses all of these protections are safeguarded.

Confidential © Aisle AI All rights reserved.

# Steps to Bringing it All Together

Training an SLM for a retailer, CPG brand, distributor, or broker involves a series of steps. First, there's the collection of first-party data, which comprises information gathered directly from the company's interactions with its customers. As mentioned, third-party data comes from external sources not directly affiliated with the company, such as market research reports, point-of-sale scan data, demographic data, social media data, and even weather data.

Once this data is gathered, the next step is to clean, process, and structure it for model training purposes. This involves tasks like removing duplicates, handling missing values, and standardizing formats to ensure consistency across the board. From there, relevant features are extracted from the data, and tailored to the specific objectives of the company. For instance, in the retail sector, these features might include customer demographics, purchase patterns, product preferences, and seasonal trends.

These features play a crucial role in helping the model grasp the underlying patterns and relationships within the data. With the data prepped and features identified, the SLM undergoes training using both the first-party and third-party data. Throughout this process, the model learns to identify patterns, correlations, and trends that align with the retailer or CPG brand's business goals, all within the confines of your own private instance.

By seamlessly integrating key data sources like billions of credit card transactions and retail point-of-sale transactions within seconds, Gen AI running on an SLM unveils spending patterns and sheds light on consumer behavior. A solution like the Aisle AI Data Intelligence Platform pinpoints the effectiveness of promotional activations down to specific store locations, facilitating data-driven decision-making at an unprecedented speed and scale. Without the computational prowess of Gen AI on SLM solutions, the sheer volume of data and the insights it can unlock would remain unattainable.

At Aisle AI, we're pioneering the training of SLM's for the CPG industry, making the power of Gen AI accessible to all within the CPG ecosystem. By training an SLM on both first-party and third-party data, retailers and CPG brands can tap into a broader spectrum of insights and perspectives, empowering better decision-making, enhancing customer experiences, and ultimately, driving business growth.

# Conclusion

In summary, this whitepaper has discussed the distinction between AI and ML, showing how ML serves as the foundational element of AI by instructing algorithms to glean insights from data. We've introduced the concept of Generative AI, or "Gen AI," which imbues new data with vitality based on accumulated knowledge.

Delving deeper, our exploration has contrasted the efficacy of Large Language Models (LLMs) versus Small Language Models (SLMs) within the CPG sector. Our analysis reveals that SLMs, finely attuned to the nuances of CPG and retail, offer more precise insights that resonate with decision-makers.

We have emphasized the pivotal role of proprietary data—such as transaction records, customer profiles, and related metrics—in creating timely actionable data-driven insights within the CPG domain. We have underscored the significance of augmenting internal data with third-party sources, such as point-of-sale scan and demographic data, to enrich our understanding and strategic decision-making. We have also discussed best practices around using “clean rooms” to safely and securely intermix these data sources to create even more robust and directly applicable business insights.

Addressing a pertinent concern, we have discussed strategies for mitigating the occurrence of AI-generated inaccuracies, commonly referred to as "AI hallucinations." Recommendations include leveraging high-quality data inputs and maintaining vigilant human oversight to ensure the integrity of AI-driven insights.

Aisle AI is here to partner with you as your AI co-pilot on your AI journey. Utilizing our commercialized, secure, military-grade technology, the Aisle AI Data Intelligence Platform seamlessly integrates data analytics and predictive intelligence within a private environment. With Aisle AI, you can rest assured that your competitors won't gain insights from your queries. Instead, you'll gain a more comprehensive understanding of the market landscape, empowering CPG businesses to not only respond to current trends but also anticipate future shifts.

In conclusion, we hope this whitepaper serves as a comprehensive guide for navigating the intersection of AI, ML, and CPG, equipping stakeholders with actionable insights to drive informed decision-making and achieve sustained success in the evolving marketplace. Together, let's leverage the power of AI to foster innovation, enhance efficiency, and fuel growth in the retail sector. The future of CPG is promising, and we're here to help you thrive.

## AUTHORS



Kevin Senn has carved a unique path that embodies innovation, dedication, and foresight. He recognized early the untapped potential of data analytics and AI in transforming the CPG landscape. Kevin is driven to transform the CPG industry with a pioneering data analytics and AI platform. His track record includes patented innovations in AI image recognition, data analytics, and CPG shelf space positional optimization.



Chris Hughes is a seasoned executive leader with deep roots in the Consumer and Retail industries. A management advisory consultant by trade, Chris has helped numerous consumer companies navigate digital disruptions and win in the market over the past 25+ years, most recently leading the Americas consumer & retail consulting and analytics practice at DXC Technologies.



George Martinez boasts over three decades of experience as a seasoned executive in the Consumer Packaged Goods (CPG) sector, particularly within the food and beverages industry. His illustrious career flourished primarily at PepsiCo Bottling, where he honed his expertise in brand management and field services, earning recognition for his contributions to the industry.



The team at Aisle AI are industry-leading CPG experts and trailblazers in data intelligence transformation, committed to harnessing the power of AI to create powerful and dynamic data and retail analytics that are simple, accessible, and affordable for all in the CPG industry. The Aisle AI Data Intelligence Platform provides a holistic solution featuring a visualization dashboard, advanced predictive modeling, and advanced data management. Through seamless integration of first and third-party data sources, the platform enables intuitive natural language search and automated anomaly detection. Real-time updates and AI-driven insights empower teams to collaborate effectively and uncover valuable trends. With Generative AI and predictive modeling, Aisle AI delivers actionable strategies for retail optimization. Military-grade security ensures data integrity and role-based access, expediting informed decision-making in CPG sales, distribution, and operations.