EVIDEN CREATES KODA, A GENERATIVE AI VISUAL KIOSK PLATFORM AT THE EDGE, USING SUPERMICRO SERVERS & NVIDIA GPUS

KODA is a Life-Like Digital Avatar powered by Edge AI, that can Greet, Assist, and Guide Customers in Real-World Retail Stores.

Figure 1 - Supermicro Servers. Top: SYS-110P, Bottom: SYS-221HE

KODA Digital Avatar (image courtesy of Eviden)

TABLE OF CONTENTS

Executive Summary

Artificial Intelligence (AI) is finding its way into every aspect of our lives, bolstered by the rapid rise of generative AI following the release of ChatGPT in 2022. Although the first industries and businesses to implement AI have been digital, other companies are now looking for novel ways to improve their businesses and serve customers using the technology. According to a 2023 MIT Technology Review Insights survey, 78% of executives agree that scaling AI/ML use cases to create business value is a top priority.[1]

Eviden – a technology company specializing in data-driven, trusted, and sustainable digital transformation– is

[1] MIT Technology Review Insights. (2023). *The Great Acceleration: CIO Perspectives on Generative AI*. Cambridge, MA: Massachusetts Institute of Technology.

pioneering an Edge AI solution using both generative and predictive AI that brings this cutting-edge technology to brick-and-mortar retailers. KODA is an AI-powered platform designed to revolutionize user interactions through state-of-the-art visual kiosk technology. The KODA platform is a digital advisor and concierge system that uses generative and predictive AI to assist customers in retail stores in finding routes to a specific product or giving product recommendations. Customers interact with KODA just like they would a store employee and can ask questions and receive answers using natural language.

Implementing AI in a physical kiosk such as KODA presents unique challenges to overcome and opportunities for those who can accomplish it. The first is in bringing an inherently digital experience to a physical setting. Shoppers have become extremely familiar with e-commerce over the past decade, but the same conveniences found online have not transferred to brick-and-mortar retailers. This is one of the goals of the KODA project. The other set of challenges to overcome in such an implementation are the technical aspects of an Edge AI project. For KODA's kiosk form factor, these involve considerations like size, weight, and power (SWaP), remote software management, energy efficiency & network connectivity. These challenges must be tackled to deploy a solution within the constraints on physical space availability, noise output, and serviceability that arise on a project like KODA.

Despite these hurdles, KODA exemplifies the transformative potential of Edge AI in creating responsive and personalized user experiences. At the core of the KODA solution's operational excellence is a seamless integration of Supermicro's robust hardware, complemented by the cutting-edge NVIDIA® AI Platform to handle software requirements, all combined thanks to Eviden's data science expertise.

KODA's Wayfinding Capabilities

KODA stands out by transforming the customer experience through personalized, intuitive interactions. This opportunity allows retailers to offer their customers tailored recommendations and services, thus significantly enhancing the overall customer journey within various environments. Imagine a customer walking into a grocery store and being greeted by Koda's digital avatar on a kiosk. The customer is asked if they have any questions, and when they ask about the beverage aisle, they are directed to that exact location with a real-time 'fly over' of the store layout.

Unlike traditional wayfinding solutions that rely on legacy WIFI or GPS technology, KODA leverages the cutting-edge capabilities of NVIDIA Omniverse™ to reimagine product-level navigation in three dimensions: X, Y, and Z. This innovative approach not only eliminates the need for planogram compliance but also enhances the accuracy and efficiency of locating products within complex and diverse environments. As seen in Figure 2, the KODA kiosk might be placed near the doors to the store, allowing customers to interact with the avatar upon entering and receive wayfinding assistance or concierge support.

February, 2024

Figure 2 - KODA Wayfinding Solutions

As retailers have increasingly replaced personnel with touchscreens and self-checkouts, customers have seen limitations imposed on their ability to deal with exceptions and questions. KODA bridges this gap by providing more customer and employee interaction and engagement flexibility.

At the core of KODA's exceptional wayfinding capabilities is its ability to integrate real-time data specific to each environment. This integration enables KODA to provide customers with immediate, accurate, and personalized directions to their desired destinations. Whether it's guiding customers to a particular product in a retail store or offering recommendations for new experiences, KODA's real-time data integration ensures a seamless and intuitive navigation experience. This capability is a testament to KODA's ability to act in the role of a model employee, consistently delivering valuable insights to improve both customer satisfaction and operational efficiency in retail and service settings.

Personalization Through KODA

In today's retail landscape, personalization is not just a preference; it's an expectation. Eviden created KODA with this in mind so that retailers can easily customize the platform alongside their offerings. From customizing the visual avatar to reflect known brand ambassadors or characters to accessing customer history for tailored recommendations, KODA ensures that each interaction is as unique as the customers themselves. This level of personalization enhances the customer experience, making every interaction with KODA feel personal and relevant.

Imagine a customer approaching KODA with a question as simple as what to bring to a barbecue. KODA's response goes beyond mere suggestions; if the customer desires, it can provide a brief roadmap for their journey through the retail space, including recommendations, store specials, and even beverage pairings. This is further augmented by a QR code that customers can scan to use Augmented Reality on their phones, guiding them through the aisles to their desired products. Figure 3 shows a few of the personalized interactions that are possible with KODA.

February, 2024
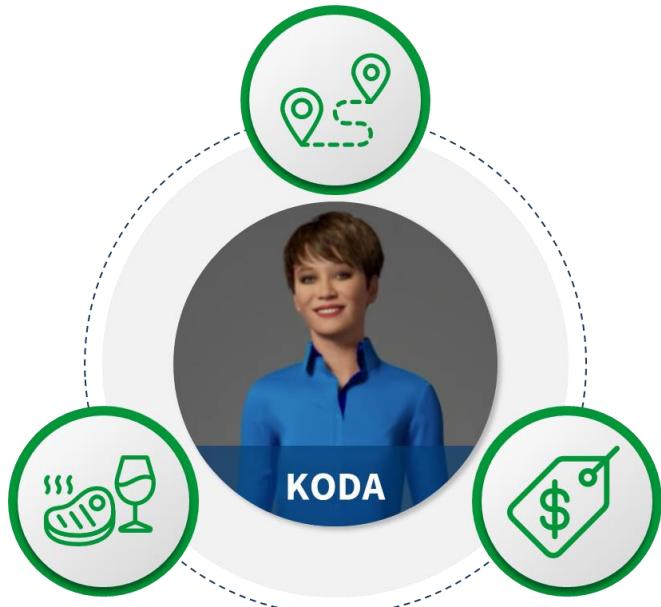
Figure 3 - Personalized Interactions Powered by KODA

To achieve cutting edge personalization, KODA is designed to leverage human factors engineering, a subcategory of psychology. This approach helps facilitate an intuitive, immersive, and interactive experience. Realism is a core emphasis. This psychological approach to engineering manifests visually in two ways. First, in the 3D digital twin environment of the store, it is generated using 3D LIDAR, 2D optical scans, and simulations of lighting of the space. Second, these principles are put into practice on KODA's character avatars, which use the latest engines used both in the film and gaming industries. As customers interact with KODA, they receive real-time responses, eye contact, and natural conversation that make them feel as if KODA is just another friendly person in the room with them.

Eviden has made great efforts to ensure that KODA is designed to interact with our diverse population, meaning the platform is extendable and expansive in its ability to cover diversity in languages, accents, ethnicities, skin tones, body shapes, and sizes, disabilities, and other physical and personal characteristics. The system is built with the capability to continuously monitor interactions and to override and correct systematic biases that may occur as a result of some of the AI models that may be used.

Operational Impact and Market Reception

KODA is a unifying platform that brings together diverse partner solutions from cloud to edge technologies. This integration capability extends across servers, cloud technology, developers, employees, customers, and various infrastructure providers, facilitating a cohesive ecosystem. Through KODA, Eviden has created a new way to improve the customer experience in retail stores and transit facilities. The system has been well received by grocery chains, convenience stores, campuses, and airports.

The market's response to KODA has been positive, with infrastructure providers particularly excited about the new dimensions of immersive interactivity that KODA introduces at the edge. This enthusiasm is driven by the potential for increased hardware and connectivity requirements and the futuristic appeal of KODA's technology. On the other hand, retailers have welcomed the opportunity to elevate their customer service to a "white glove" level, distinguishing their offerings from traditional wayfinding

February, 2024

solutions. The tangible benefits of KODA's deployment, reflected in enhanced customer satisfaction and operational efficiencies, underscore its value proposition and the transformative impact it has on the market.

Synergy with Supermicro and NVIDIA

The collaboration between Eviden, Supermicro, and NVIDIA has been vital to enhancing KODA's capabilities, particularly in terms of performance, reliability, and scalability. This partnership has leveraged the strengths of each entity, with Supermicro providing the robust hardware infrastructure alongside advanced technologies like the NVIDIA Omniverse platform, which has been pivotal in redefining wayfinding solutions in three dimensions.

> *By integrating Supermicro's GPU-based servers, KODA kiosks leverage the power of large language models (LLMs) and real-time inference, enabling swift, intelligent processing that elevates the quality of customer interactions. "* - Jenny Jacob Senior Vice President & Head of Industry

The KODA platform utilizes two separate Supermicro systems with onboard NVIDIA GPUs. The first is a Supermicro SYS-110P with an NVIDIA L40 that lives inside the kiosk and handles all of the sentiment analysis (through computer vision), NVIDIA Riva ASR (Automatic Speech Recognition) for response from the LLM, and rendering of the avatar with the help of NVIDIA Audio2Face. The 110P system is used here because it addresses the size and airflow requirements necessitated from being inside the kiosk. The second system is a Supermicro Hyper E, SYS-221HE, with dual NVIDIA A6000 Ada GPUs. This system is located outside of the kiosk and is the host for the LLM. Conveniently for retailers, it is possible for multiple KODA-based kiosks to run on this single external LLM hosting system, aligning with Eviden's vision that KODA be integrated at points throughout a retail store.

Kiosk Operation Server

SYS-110P-20C-FRAN8TP

Dimensions (mm)

H	W	D
43.0 X	437.0 X	399.0

Key Features

Single L40

LLM Host Server

SYS-221HE-TNR(D)

Dimensions (mm)

H	W	D
88.9 X	436.9 X	574.0

Key Features

Dual RTX 6000 Ada

February, 2024

Products such as KODA, built with Edge AI in mind, require specialized hardware and software for optimal performance. Supermicro and NVIDIA's collaborative approach provides just that—a comprehensive full-stack solution that integrates CPUs, GPUs, optimized memory, and the NVIDIA AI Enterprise software platform, all orchestrated within the resilient infrastructure of Supermicro's platforms. Real-time video and audio processing, generative AI output, and management of large data sets about store layouts and products require robust processing capabilities. Similarly, physical hardware must be highly reliable because these kiosks might be moved, bumped, or spilled on during business hours. Supermicro's systems provide that resiliency while also providing energy efficiency, noise reduction, and security in physical locking and digital data protection.

Focus on Privacy and Security

In the era of digital transformation, data privacy and security are paramount, especially for technologies that interact closely with users, like KODA. Eviden takes these concerns seriously, adopting a robust framework to safeguard user data. KODA is designed to operate within the strict confines of Europe's stringent privacy laws, setting a high standard for data protection. The platform's approach to privacy is proactive and preventive; KODA never stores individual data, ensuring that personal information is not at risk of unauthorized access or misuse.

Eviden, a global leader in cybersecurity, is committed to creating solutions that comply with the General Data Protection Regulation (GDPR) following European regulations. KODA's architecture is built to respect user privacy at every interaction point. While KODA can be trained to offer personalized services, it does so without directly accessing or storing user data. This distinction is crucial, as it allows KODA to provide tailored recommendations and services based on retail data provided by customers for training AI models without compromising individual privacy.

In terms of security, KODA can have a positive impact on the locations where it is situated, thanks to integration with Eviden's Ipsotek video surveillance software. This technology can be used in airports or campuses to monitor public safety threats and help security teams stay informed. At grocery or convenience stores, KODA's Ipsotek integration can help loss prevention teams detect situations that require a response, contributing to real-time prevention of shrinkage.

Plans for the Future

Eviden is at the forefront of the evolving landscape of retail technology, where the integration of digital and physical experiences is increasingly paramount. The future of retail technology, as envisioned by Eviden, involves a seamless blend of online convenience and in-store interactivity, where KODA's capabilities in enhancing customer journeys and providing detailed analytics play a crucial role. Unlike online shopping, where data is readily available by nature, these insights have been lacking in brick-and-mortar settings. KODA addresses that opportunity by allowing retailers to see how customers are shopping and engaging at scale and in real time.

Looking ahead, Eviden is committed to continuous innovation with KODA, focusing on expanding its realism and functionality to deliver even more immersive experiences. The roadmap includes enhancements in augmented reality (AR) and virtual reality (VR) technologies, offering customers new ways to interact with products and services in a virtual space. Additionally, Eviden plans to introduce more seamless mobile options, further integrating KODA into the everyday lives of consumers.

Powering Up Technology Adoption

February, 2024

What should retailers consider when looking to adopt Edge AI solutions like KODA? Eviden suggests starting with a pilot near the entrance of a store, as it allows customers to engage with the technology from the moment they enter, thereby maximizing exposure and early adoption benefits. This positioning also facilitates the collection of real-time customer feedback and interaction data, essential for continuous improvement and customization of the technology to meet specific retail needs. By implementing KODA at key customer entry points, retailers can quickly demonstrate its value and convenience, encouraging widespread acceptance and use among their clientele.

For successful integration of digital advisor systems like KODA, it is important to focus on building a solid and focused data foundation. This foundation should integrate seamlessly with all operational aspects of the retail space and directly address real customer challenges. Starting with a strong data foundation ensures that the technology enhances the customer experience and contributes to the overall efficiency and effectiveness of retail operations. Retailers can then expand the implementation of KODA's capabilities into other parts of their stores.

Summary

The implementation of Eviden's KODA platform has set a new benchmark in the digital advisor and concierge space, addressing critical challenges in retail and service environments. Utilizing Edge AI in this way allows retailers to address these challenges and enhance customer satisfaction, increase sales through personalized recommendations, and improve operational insights. Whether suggesting a wine pairing for a steak dinner or routing a parent and child to their favorite snack food, KODA's digital avatar is available at a moment's notice with perfect assistance.

This type of real-time inferencing, using generative and predictive AI, would not be possible without Edge AI. Data transmission costs are too high, and speeds are too slow, resulting in an experience that lags for crucial seconds after every customer query. By handling processing on-site and securing that data with physical and digital measures, the collaborative Supermicro and NVIDIA solution makes KODA possible. Furthermore, Supermicro's flexible range of solutions ensures that retailers utilizing the KODA platform can scale up their implementation as much as needed. Whatever the requirements, solutions are available to expand memory, processing power, and storage to meet any situation.

Beyond retail, Edge AI projects hold significant promise in areas such as healthcare, where AI-driven diagnostics and patient management systems can improve care delivery; in manufacturing, where predictive maintenance and quality control can enhance operational efficiency; and in smart cities, where AI can optimize everything from traffic management to public safety. The possible areas of ROI for businesses implementing Edge AI, powered by Supermicro's hardware and NVIDIA's software, include reduced operational costs, increased revenue opportunities, and enhanced customer experiences.

The journey of KODA from concept to production underscores the transformative power of Edge AI. As Eviden continues to innovate and expand KODA's capabilities, the future looks promising for businesses ready to embrace the new capabilities of AI.

February, 2024

For More Information

Supermicro Edge AI Solutions - www.supermicro.com/edge-ai

Supermicro SYS-110P - https://www.supermicro.com/en/products/system/iot/1u/sys-110p-frn2t

Supermicro SYS-221HE - https://www.supermicro.com/en/products/system/iot/2u/sys-221he-ftnrd

Eviden Generative AI - https://eviden.com/solutions/generative-ai

SUPERMICRO

As a global leader in high performance, high efficiency server technology and innovation, we develop and provide end-to-end green computing solutions to the data center, cloud computing, enterprise IT, big data, HPC, and embedded markets. Our Building Block Solutions® approach allows us to provide a broad range of SKUs, and enables us to build and deliver application-optimized solutions based upon your requirements. More information at https://supermicro.com

NVIDIA

Since its founding in 1993, NVIDIA (NASDAQ: NVDA) has been a pioneer in accelerated computing. The company's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, ignited the era of modern AI and is fueling industrial digitalization across markets. NVIDIA is now a full-stack computing company with data-center-scale offerings that are reshaping industry. More information at https://nvidianews.nvidia.com.

EVIDEN

Eviden is a next-gen technology leader in data-driven, trusted and sustainable digital transformation with a strong portfolio of patented technologies. With worldwide leading positions in advanced computing, security, AI, cloud and digital platforms, it provides deep expertise for all industries in more than 47 countries. Bringing together 53,000 world-class talents, Eviden expands the possibilities of data and technology across the digital continuum, now and for generations to come. More information at https://eviden.com.

February, 2024