PRESIDIO® | CASE STUDY

HIGH-TECH

MODERNIZATION

Leading Pioneer in Digital Design Solutions

The Customer

The customer is an American multinational software corporation that makes software products and services for the architecture, engineering, construction, manufacturing, media, education, and entertainment industries. The chatbot assistant is used to interact with customers and to resolve the query raised by them. The customer became best known for developing a broad range of software for design, engineering, and entertainment—and a line of software for consumers. The manufacturing industry uses customer's digital prototyping software to visualize, simulate, and analyse real-world performance using a digital model in the design process. The customer is best known for a wide range of software products for design, engineering, and entertainment, including digital prototyping software for the manufacturing industry, building information modeling software for construction planning, and software for media and entertainment purposes such as visual effects, color grading, editing, animation, game development, and design visualization.

The Challenge

To enhance user interactions with their chatbot. The goal was to significantly improve the customer experience by empowering the chatbot to effectively handle a wide range of user queries without requiring human intervention.

Why AWS

Customers selected AWS as the preferred cloud service provider for building secure and scalable enterprise applications, particularly from the perspective of machine learning (ML). AWS offers a comprehensive set of ML services that are well-integrated with their broader cloud ecosystem, making it an ideal choice for Customer's ML needs.

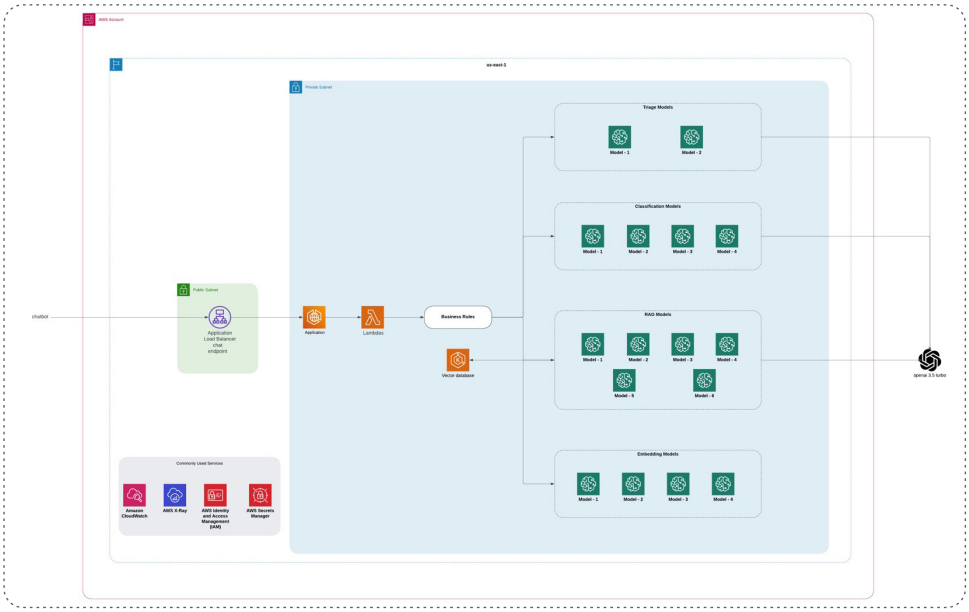
One of the key services used by the solution is AWS SageMaker, which provides a fully managed platform for building, training, and deploying ML models at scale. SageMaker's frameworks simplify the ML development process, allowing the solution to focus on model experimentation and deployment.

The Solution

To achieve the chatbot's goal, an AI-driven architecture was implemented, featuring the GenAI prompt method and leveraging embedding and vector databases. This architecture comprises several layers, each serving a distinct function:

1. **Triage Layer:** This foundational layer is powered by a model deployed on a SageMaker endpoint, employing the OpenAI prompt method to swiftly categorize incoming queries as either domain-related or non-related, setting the stage for efficient handling.
2. **Classification Layer:** Building upon the triage layer, multiple specialized models deployed on SageMaker endpoints are strategically invoked, following predefined business use case flows. These models employ OpenAI prompt techniques to accurately classify queries into specific workflows for enhancing contextual understanding and response accuracy.
3. **RAG Layer:** At this Retrieval-augmented generation layer, additional models are deployed on SageMaker endpoints which dynamically rephrase queries based on user context and leverage the power of the vector database and embedding models. Here, the Weaviate vector database plays a pivotal role, assessing similarity scores and retrieving the most relevant responses from preloaded FAQs, ensuring a seamless and precise interaction experience.
4. **Embedding Layer:** Integral to the RAG layer, these open-source Hugging Face models, deployed as Flask applications on SageMaker endpoints, further enhance response quality and relevance by harnessing advanced embedding techniques.

High-level Architecture



About Presidio

Presidio is a global digital services and solutions provider accelerating business transformation through secured technology modernization. Highly skilled teams of engineers and solutions architects with deep expertise across cloud, security, networking, and modern data center infrastructure help customers acquire, deploy, and operate technology that delivers impactful business outcomes. Presidio is a trusted strategic advisor with a flexible full life cycle model of professional, managed, and support and staffing services to help execute, secure, operationalize, and maintain technology solutions. For more information, visit www.presidio.com.

Business Outcome

The implementation of this architecture has resulted in a substantial reduction in agent interactions, with the chatbot now handling as much as 65% of user queries autonomously. This enhancement has not only improved the overall user experience but has also positively impacted business outcomes.

Technologies Used

AWS SageMaker, Lambda, DynamoDB, S3, API Gateway, Secrets Manager, VPC, Route 53, ECR, ECS, EKS, and Redis.

Partners

aws

For more information on how we connect IT of today to IT of tomorrow, visit presidio.com