

# ShopSmart

## Prompt Engineering - Final Project

Sai Durga Mahesh Bandaru - 07-16-2024

**ShopSmart is an intelligent e-commerce platform designed to enhance users shopping experience using Generative AI**

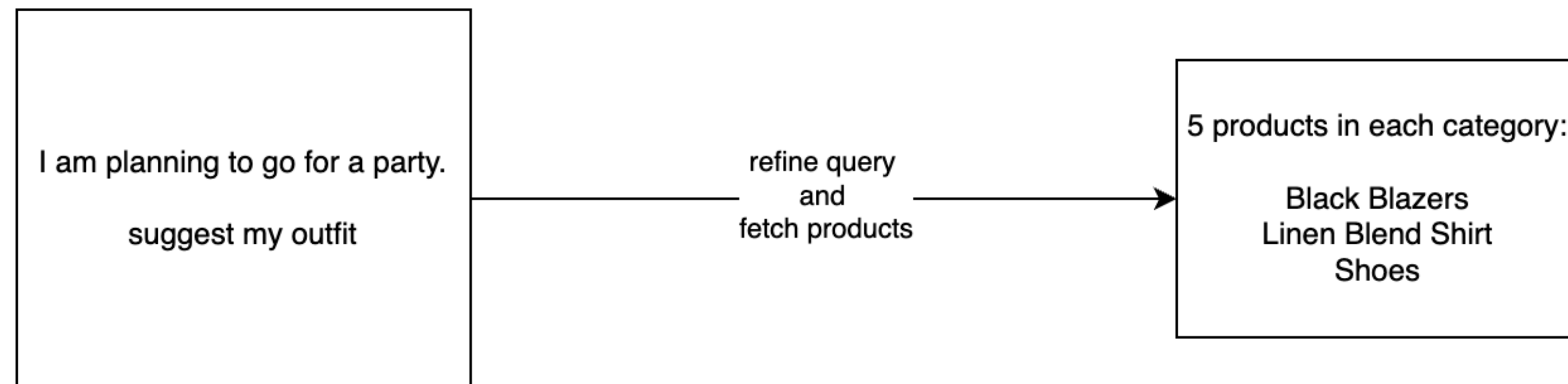
# Objectives

- AI-powered product question answering system
- Leverages user reviews to provide insights
- Enhancing product research and buying decisions
- Simplify product research process for online shoppers
- Provide accurate answers to product questions using real user experiences
- Improve shopping decisions through aggregated user knowledge

# How ShopSmart Works

## Search in ShopSmart

- User can ask what he needs .System analyses the query to understand the information need

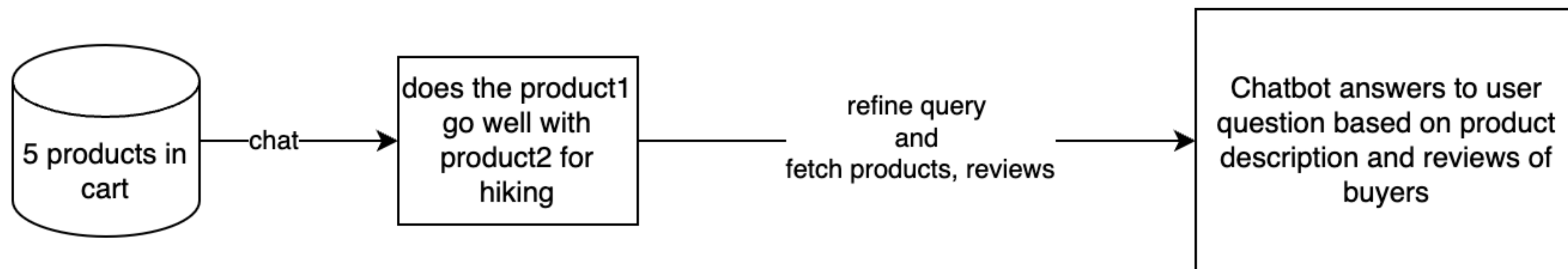


- User Can add products that he liked into cart

# How ShopSmart Works

## Cart in ShopSmart

- User can ask questions about products in cart before buying. All relevant user reviews are retrieved and added into context for answering user queries.



# Project Description

ShopSmart is an innovative AI-powered system designed to revolutionize the online shopping experience by providing intelligent answers to product-related questions using real user reviews. The system integrates advanced natural language processing techniques with a vast database of user-generated content to offer shoppers quick, accurate, and relevant information about products they're interested in purchasing.

# Benefits for Shoppers

- Access to real user experiences and opinions
- Quick answers to specific product questions
- Reduced time spent reading through numerous reviews
- More informed purchasing decisions
- Increased confidence in product selection

# Architecture Overview

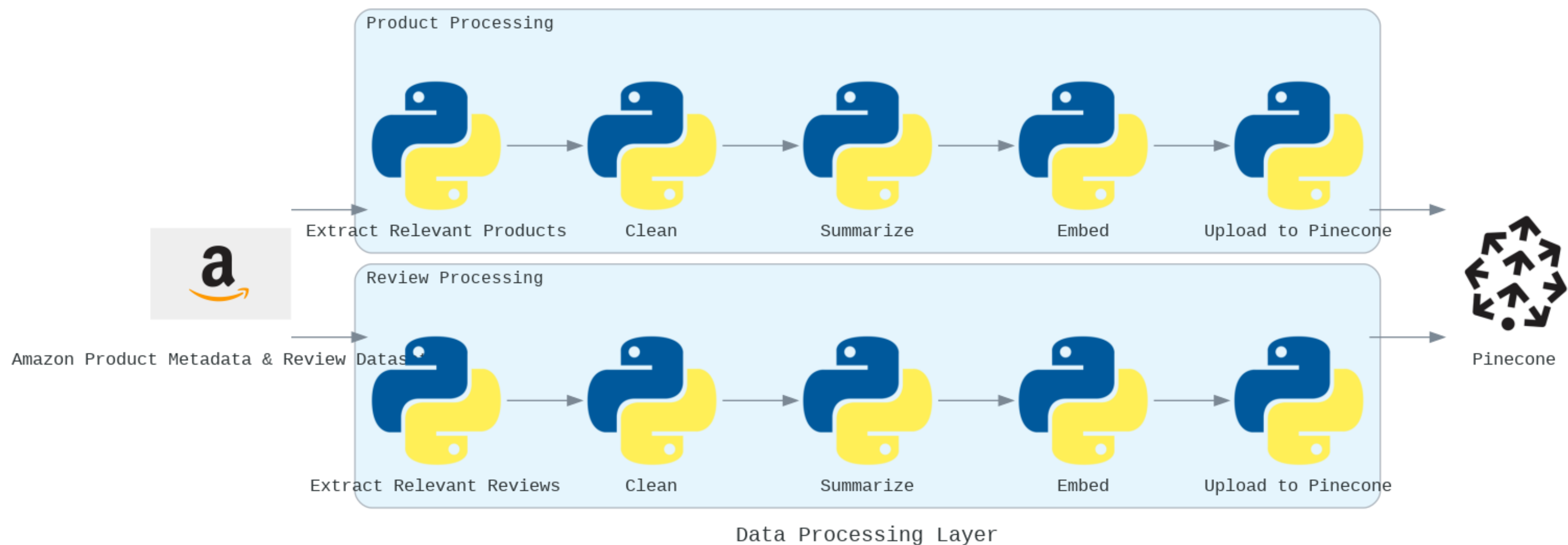
## Key Components

- Frontend: Streamlit for user-friendly interface
- Backend: FastAPI for efficient query processing
- Database: Snowflake for storing and accessing review data
- AI Integration: OpenAI for natural language understanding and response generation

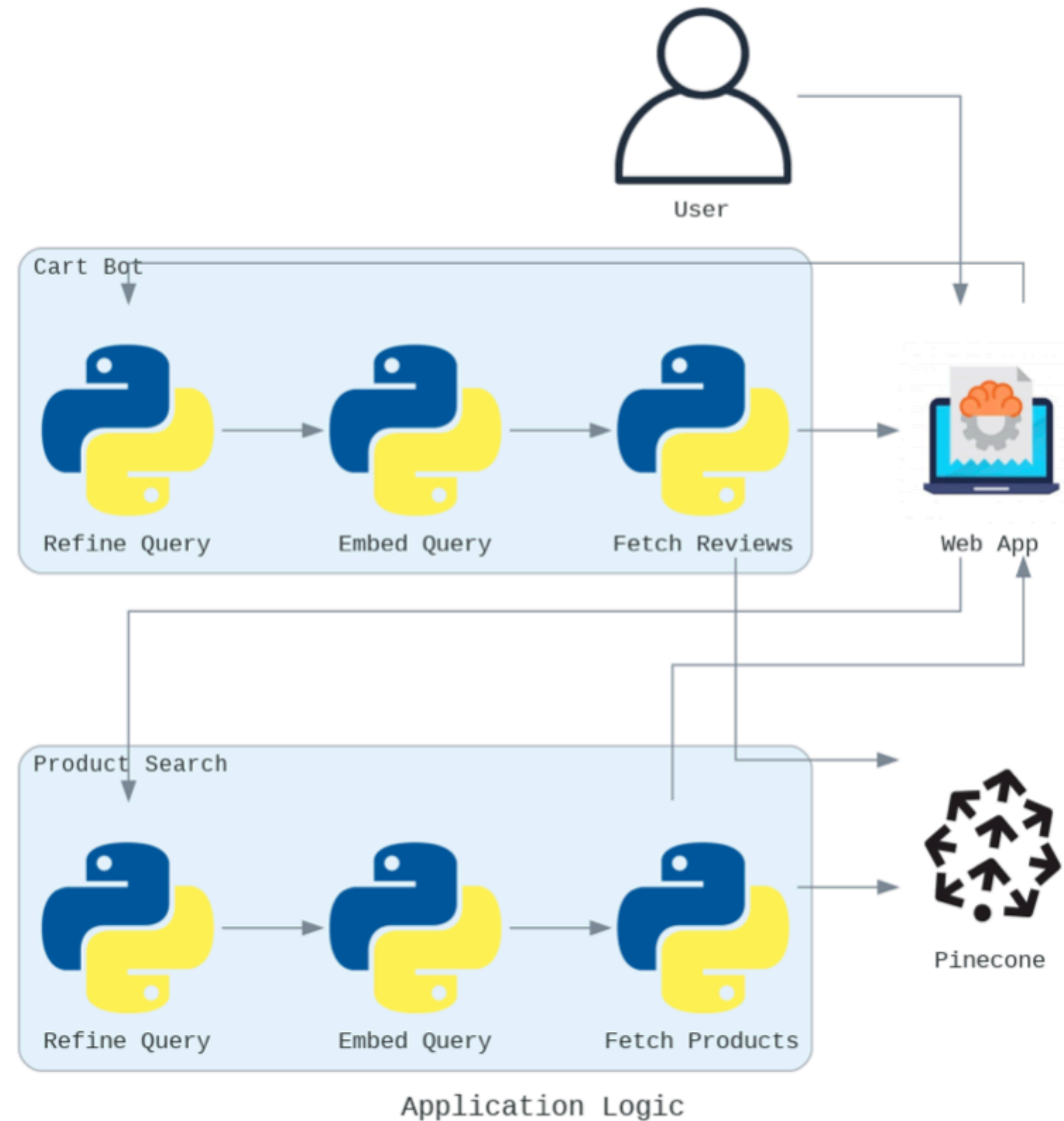


# Architecture

## Data Processing



# Application Flow Architecture



# RAG Implementation

## Data processing

- **Extract**

- Collect and clean information of product from 10 different categories in amazon scrapped dataset
- Collect and clean user reviews of all the above products

- **Clean**

- Preprocess the review and product info (tokenization, removing stop words, etc.)
- Clean the reviews to remove offensive words without losing context

# **RAG Implementation**

## **Embedding Generation and Indexing**

- **Summarise**
  - Summarise the product info of each product
  - Summarise the reviews in different chunks for each product
- **Embed and Upload to pinecone**
  - Create dense vector representations (embeddings) for each product and review chunk
  - Use a pre-trained language model (e.g., BERT, RoBERTa) for embedding generation
  - Store embeddings alongside the metadata in pinecone

# RAG Implementation

## Query Processing

- **Search**

- Identify Category user is referring to with Text Classification model
- Develop a prompt to sequentially identify and analyze key query aspects (features, occasions, sub-queries) using a step-by-step reasoning process.
- Embed each refined Sub-Query with same embedding model used in data processing

- **Cart-Bot**

- Develop a prompt to identify key aspects of the query (e.g., products user is referring to based on chat history, specific features, occasions)

# RAG Implementation

## Retrieval System

- **Search**

- Use embedding of subqueries to retrieve products from pinecone.
- Experiment with different similarity matching techniques to find balance between accuracy and latency
- Return the retrieved products for each sub-query

- **CartBot**

- Extract summaries of reviews matching user question from Pinecone for each product in cart
- Extract product summaries for each product in cart

# RAG Implementation

## Context Preparation

- **Search**
  - Use product info/summary for all the products fetched in each subquery
- **Cart Bot**
  - Use product summaries and matched review chunks for each product in cart.

# RAG Implementation

## Answer Generation

- **Search**

- Develop prompts that effectively combine the query, retrieved context, and desired output format to present the retrieved products for each subquery and create sales text that convinces the user that the given products match their needs for the occasion

- **CartBot**

- Develop prompts that effectively combine the query, retrieved context, and desired output format to answer user questions about products in cart



# Testing

## **Review Relevance Precision:**

- How accurately the retrieved reviews match the specific aspects of the user's product query.

## **Review Coverage Recall:**

- The ability to retrieve a comprehensive set of reviews that cover all aspects mentioned in the user's query.

## **Answer Relevance:**

- How well the generated answer addresses the user's specific question using the information from the retrieved reviews.

## **Noise Robustness:**

- The system's ability to handle unclear, poorly worded, or off-topic queries without a significant decrease in answer quality.

# Testing

## **Answer Consistency:**

- The degree to which ShopSmart provides consistent answers to similar queries across different sessions or users.

## **Answer Completeness:**

- The extent to which the generated answer provides all necessary information to fully address the user's query.

## **Factual Accuracy:**

- The correctness of factual information presented in the generated answers based on the available review data.

## **Answer Coherence:**

- The logical flow and overall readability of the generated answer, ensuring it's well-structured and easy to understand.

# ShopSmart Evaluation Methodology

## Test Set

- 1000 diverse product queries across 10 categories
- 10,000 associated user reviews (1000 per category)

Metric	Automated	Manual	Frequency
Review Relevance & Coverage	✓ (NLP model)	-	All queries
Answer Relevance & Completeness	✓ (BERT-based)	✓	20% of queries
Consistency & Coherence	✓ (Language model)	✓	20% of queries
Factual Accuracy	✓ (Cross-reference)	✓	25% of queries
Noise Robustness	✓ (Perturbed queries)	-	25% of queries
Information Synthesis	-	✓ (Experts)	10% of queries

# Future Work

## **Multi-modal Analysis:**

- Incorporate image and video review content alongside text
- Develop AI models to extract insights from visual product reviews

## **Voice Interface:**

- Integrate speech recognition for voice-based queries
- Develop a conversational AI interface for more natural interactions

## **Comparative Analysis:**

- Enhance the system to handle comparative queries between multiple products
- Provide side-by-side feature comparisons based on user reviews

# Key Takeaways

- There's vast potential for AI in simplifying complex e-commerce decisions
- Leveraging AI to transform raw user reviews into valuable shopping guidance.
- Developing a robust RAG pipeline capable of handling diverse product categories and high query volumes.