



Multimodal Conversational AI for Amazon

Team: Nexus Intelligence

Daichi Ishikawa
Liang Gong
Haoran Zheng

Live UI Link: <https://multimodal-conversational-ai-e-commerce.streamlit.app/>

GitHub Link: <https://github.com/HaoranZhengRaul/Multimodal-Conversational-AI-E-commerce/tree/main>

December 2024

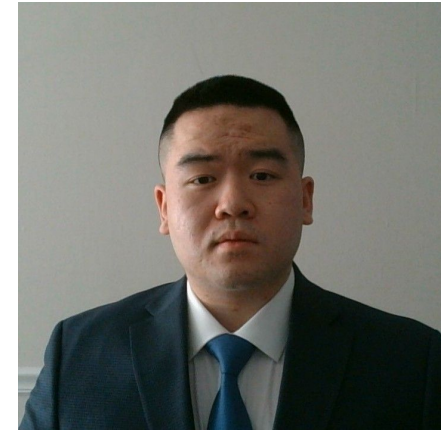
Our Team



Daichi Ishikawa



Liang Gong



Haoran Zheng

Agenda

- 1. Business Problem
- 2. Main Architecture
- 3. Data Cleaning
- 4. Embedding Generation
- 5. Vector Database
- 6. LLM and Prompts
- 7. Live Demo
- 8. Evaluation
- 9. Challenges Faced and Solutions Developed
- 10. Future Work

1. Business Problem

Current Problem

Traditional e-commerce support systems struggle to provide comprehensive product information across text and visual modes, limiting customer understanding and satisfaction.

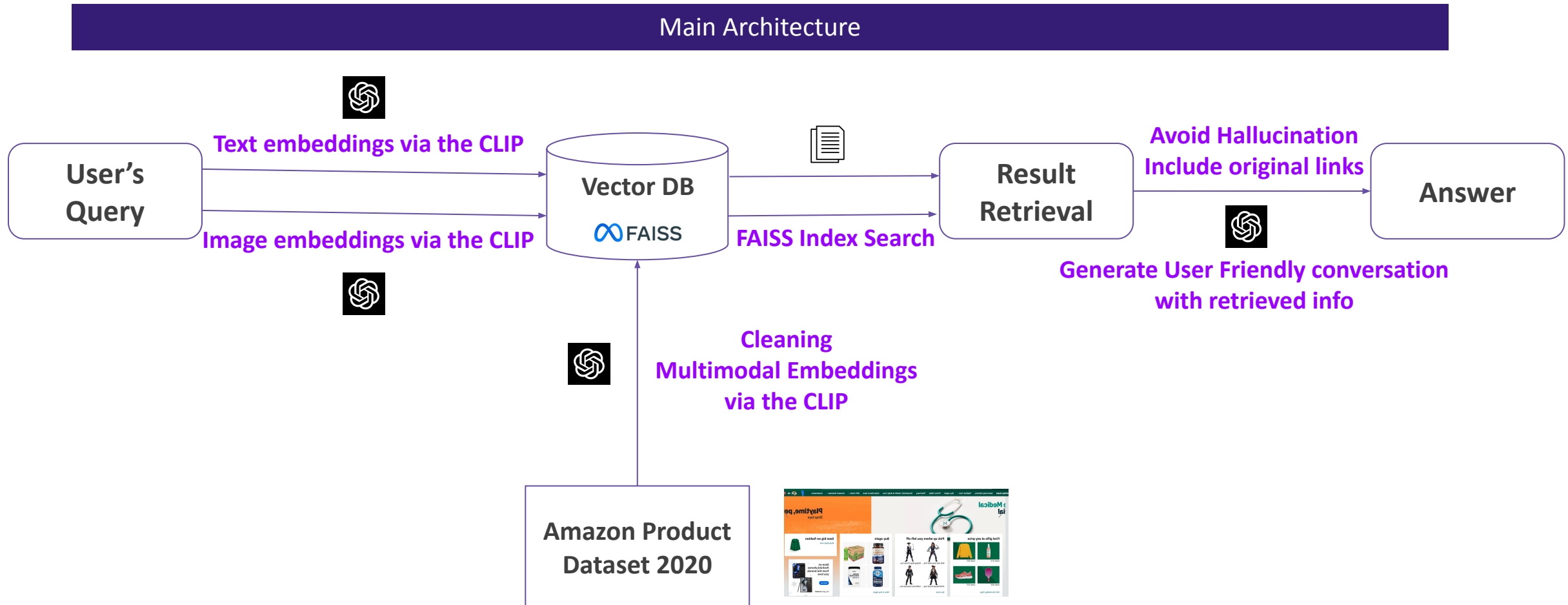


Expected Outcome

A multimodal AI chatbot that seamlessly interprets text and image queries, delivering accurate, instant product insights and transforming customer support efficiency.

2. Main Architecture

- Created text and image embeddings using CLIP
- Utilized FAISS as the vector database
- Generated the final response using GPT-4



3. Data Cleaning/Preparation

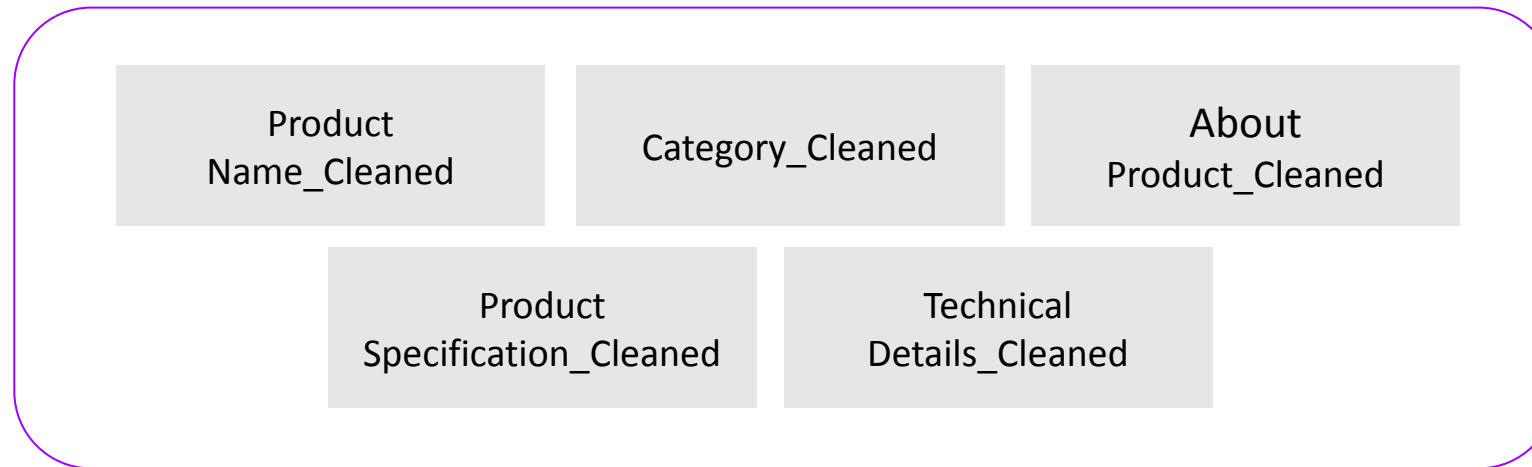
- **Columns with excessive null values:** identified and dropped.(Marketplace, Shipping Option, Free Shipping, Fulfillment, Number of Offers, Discount Percentage, Item Condition, Merchant, Delivery Time, Product Variants, Search Keywords, Specifications, Technical Details)
- **Cleaning Selling Price:** Extracts numeric values from the Selling Price column using a regex pattern and converts them to floats. Missing or invalid entries are replaced with NA.
- **Cleaning Shipping Weight:** Handles entries in Shipping Weight by: 1)Extracting numeric values for entries with "pounds", 2) Converting values from "ounces" to pounds, 3) Invalid entries are replaced with NA.
- **Converts values in the Is Amazon Seller column:** 1) Y is mapped to True, 2)N is mapped to False,3)Invalid entries are replaced with NA.
- **Cleaning Text Attributes:** The columns Product Name, Category, About Product, Product Specification, and Technical Details are cleaned by: 1)Converting text to lowercase. 2) Removing newline characters and pipes (|). 3) Stripping leading and trailing whitespace.
- **Validating Image URLs:** Ensures that all entries in the Image column start with http.

	Product Name_Cleaned	Category_Cleaned	Selling Price_Cleaned	About Product_Cleaned	Product Specification_Cleaned	Technical Details_Cleaned	Shipping Weight_Cleaned	Image	Product Url	Is Amazon Seller_Cleaned
0	db longboards coreflex crossbow 41" bamboo fib...	sports & outdoors outdoor recreation skate...	237.68	make sure this fits by entering your model num...	shipping weight: 10.7 pounds (view shipping ra...	NaN	10.7	https://images-na.ssl-images-amazon.com/images...	https://www.amazon.com/DB-Longboards-CoreFlex-...	True
1	electronic snap circuits mini kits classpack, ...	toys & games learning & education science ...	99.95	make sure this fits by entering your model num...	product dimensions: 14.7 x 11.1 x 10.2...	the snap circuits mini kits classpack provides...	4.0	https://images-na.ssl-images-amazon.com/images...	https://www.amazon.com/Electronic-Circuits-Cla...	True
2	3doodler create flexy 3d printing filament ref...	toys & games arts & crafts craft kits	34.99	make sure this fits by entering your model num...	productdimensions:10.3x3.4x0.8inches itemweigh...	show up to 2 reviews by default no longer are ...	0.8	https://images-na.ssl-images-amazon.com/images...	https://www.amazon.com/3Doodler-Plastic-Innova...	True
3	guillow airplane design studio with travel cas...	toys & games hobbies models & model kits ...	28.91	make 8 different planes at one time. experim...	productdimensions:3.5x6.2x13inches itemweight:...	go to your orders and start the return select ...	0.84	https://images-na.ssl-images-amazon.com/images...	https://www.amazon.com/Guillow-Airplane-Design...	True
4	woodstock-collage 500 pc puzzle	toys & games puzzles jigsaw puzzles	17.49	make sure this fits by entering your model num...	productdimensions:1.9x8x10inches itemweight:13...	show up to 2 reviews by default 100% officiall...	0.84	https://images-na.ssl-images-amazon.com/images...	https://www.amazon.com/Woodstock-Collage-500-p...	True

3. Data Cleaning/Preparation

- The following cleaned columns are selected for downstream processing: (Product Name_Cleaned, Category_Cleaned, Selling Price_Cleaned, About Product_Cleaned, Product Specification_Cleaned, Technical Details_Cleaned, Shipping Weight_Cleaned, Image, Product Url, Is Amazon Seller_Cleaned).
- To prepare for embedding generation, the following cleaned columns are combined to create a comprehensive product description. Including a variety of information helps improve the accuracy of subsequent searches.

Combined_Text



4. Embedding Generation - Model Architectures

1. Embedding Generation

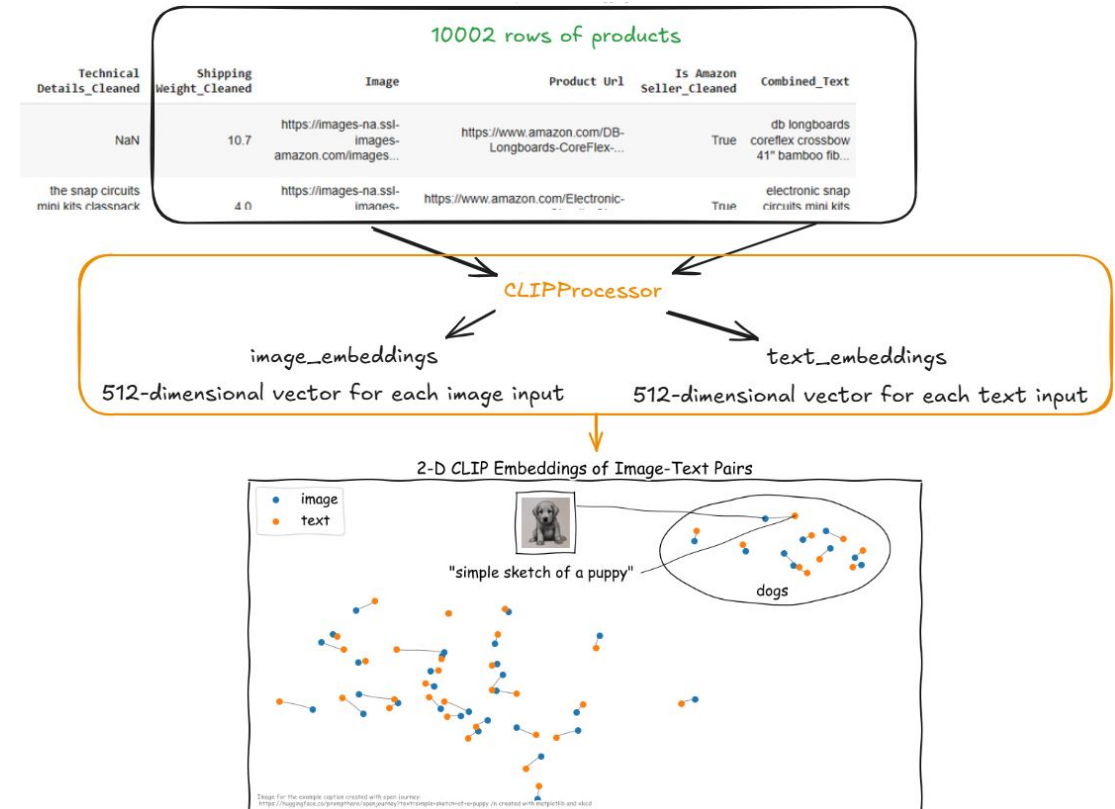
- Text and image data are converted to embeddings:
 - text_embeddings (10002, 512)
 - image_embeddings (10002, 512)
- Combined embeddings are used to retrieve the correct items.

2. CLIP Model Architecture

- Combines a transformer-based text encoder and a vision transformer (ViT)-based image encoder.
- Aligns text and image embeddings into a shared space for effective multimodal understanding.

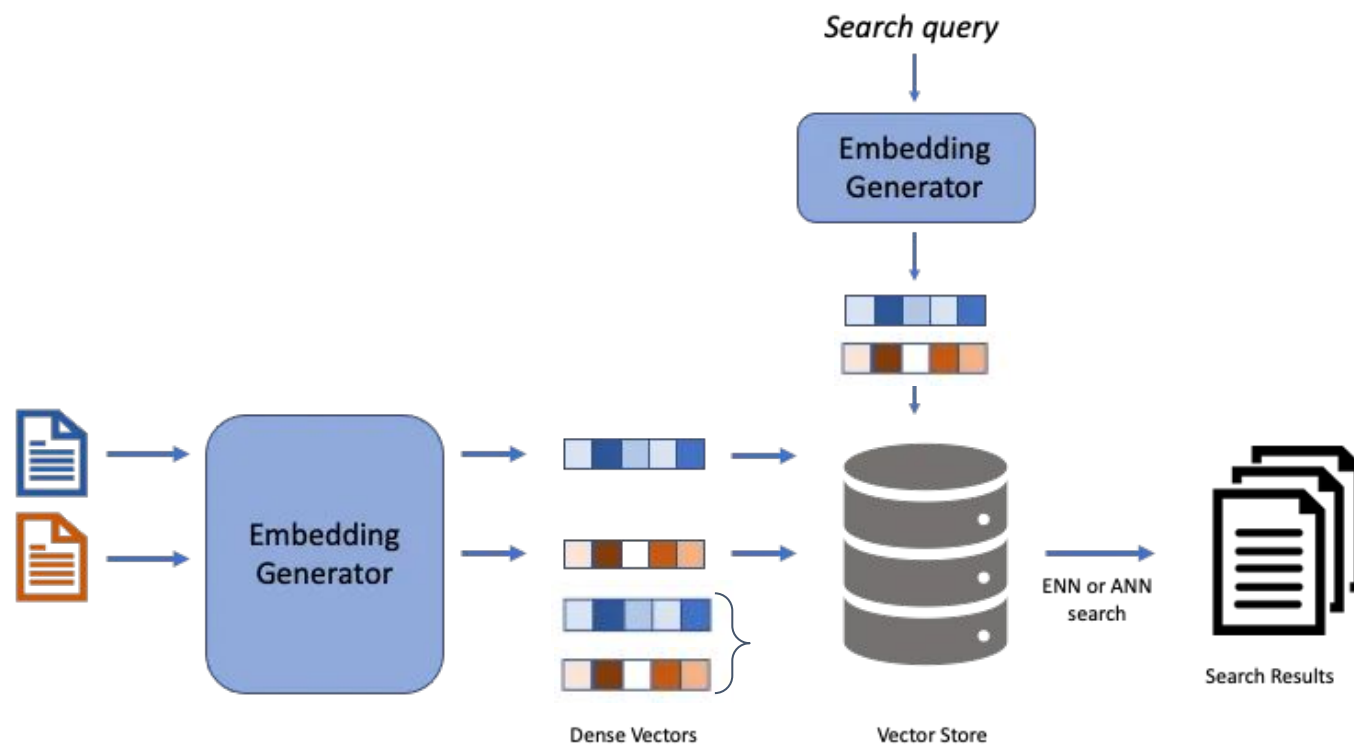
3. Key Features

- Pretrained Backbone: openai/clip-vit-base-patch32 for multimodal
- Projection Layer: Converts embeddings into a fixed 512-dimensional space.
- Fine-Grain Compatibility: Enables direct comparison of text and image embeddings with minimal preprocessing.



5. Vector Database - FAISS

- We utilized FAISS, an open-source and highly efficient library for similarity search and clustering of dense vectors.
- Syntax details:
 - wrong vs correct
 - `indices, _ = text_index.search(query_embedding, k=k)` - wrong
 - `distances, indices = text_index.search(query_embedding, k=k)` - correct
 - `text_index.search` should return distances first and then index



6. LLM and Prompts(Before Improvement)

- Tested the query "Show me some longboards."
- Generated responses using the retrieved product name, but due to the limited information available, it was unable to provide sufficient details.

```
# Create context from retrieved items
context = "\n".join(
    [f"- Product {i+1}: {row['Product Name_Cleaned']}" for i, row in retrieved_items.iterrows()]
)

# Create GPT-4 prompt
prompt = f"""
You are a knowledgeable assistant helping users with product recommendations.

Context:
{context}

Question: {query}
Answer:
"""
```

"1. "DB Longboards CoreFlex Crossbow 41" Bamboo Fiberglass Longboard Complete": This one is perfect for carving and cruising. It features a cambered platform with variable edge concave.

2. "Landyachtz Switchblade 40" Maple Longboard Complete": This longboard is highly recommended due to its drop-down design and symmetrical shape.

3. "Rayne Whip 41" Peacock Longboard Complete": This longboard is known for its flexibility and is great for freeriding and carving.

4. "Loaded Boards Dervish Sama Bamboo Longboard Skateboard Complete": This one has a drop-through truck mount style which makes it perfect for long-distance pushing, pumping and electric longboarding.

5. "Sector 9 Fractal 36" Maple Drop Through Longboard Complete": It's one of the best options due to its 8 ply maple construction which ensures durability.

Remember to consider your level of experience and specific needs when choosing a."

6. LLM and Prompts(After Improvement)

- By including details such as Product Name, Product Description, Category, and Price, we successfully retrieved more detailed and accurate information, improving the precision of the responses.
- Adjusted the prompt to include links in the responses, allowing users to access the details.

```
# Create a structured context from retrieved items
context = "\n".join(
    [
        f"{i+1}. {row['Product Name_Cleaned']}: {row['About Product_Cleaned']} "
        f"(Category: {row['Category_Cleaned']}, Price: {row['Selling Price_Cleaned']}, "
        f"Image URL: {row['Image']})"
        for i, row in retrieved_items.iterrows()
    ]
)

# Create GPT-4 prompt
prompt = f"""
You are a helpful assistant recommending products based on user queries.

Context:
Here are some products relevant to the user's query:
{context}

Based on the query, recommend the most suitable product(s) and explain why they meet the user's needs.
If the query asks for a specific product image, include the image URL in the response.

Question: {query}
Answer:
"""
```

GPT-4 Response

Based on your query, I recommend the Swagskate NG2 A.I.-Powered Electric Longboard. This longboard is not just a traditional one, it's electric and powered by A.I. which makes it unique and perfect for both beginners and experienced riders.

The longboard has hands-free control and a 900W dual-hub brushless motor. You can control it just like a traditional cruiser skateboard, by shifting your weight on the deck, with no remote required. It also comes with a remote for those who prefer it, which can control speed modes and acceleration.

The board is one of the fastest electric longboards on the market, with the ability to cruise up to 18 mph. It can also easily conquer 15-degree inclines thanks to its powerful pair of brushless rear 450-watt motors.

The deck is built for flexibility and strength, made from 6 layers of premium maple wood sandwiched between 2 layers of bamboo. It can support up to a 220 lb. max weight. The product is also backed by a 90-day limited warranty and a US-based support team.

Here are some images of the longboard:

- [Image 1](#)
- [Image 2](#)
- [Image]

swagskate ng2 a.i.-powered electric longboard with hands-free control & 900w dual-hub brushless motor 38" electric skateboard w/remote & 3.5" x 52mm pu wheels

Category: sports & outdoors outdoor recreation skates, skateboards & scooters skateboarding stanc skateboards & longboards standard skateboards

Price: 399.99

[Product Link](#)



7. LIVE DEMO



8. Evaluation

- Overall, accurate responses were achieved for text-based queries, image-based queries, and requests for specific product images.
- From the results below, Accuracy: 87.5%

	Questions	Evaluation
1	What are the features of the DB Longboards CoreFlex Crossbow 41? (Text-Based Questions)	Correct
2	Can you compare the DB Longboards CoreFlex Crossbow 41 with swagskate ng2 a.i.-powered electric longboard? (Text-Based Questions)	Half Correct
3	Can you identify the product in this image and describe its usage? (Image of a Pokemon Card) (Image-Based Questions)	Correct
4	Can you show me a picture of crossbow longboard? (Request for a Product Image)	Correct

9. Challenges Faced and Solutions Developed

Challenges

- **Data Preprocessing:** Handling missing values and malformed URLs required careful attention to maintain embedding consistency. Ensuring text concatenation preserved semantic meaning without introducing noise.
- **Scalability:** Embedding computation for larger datasets may lead to memory constraints, especially for images.
- **Index Loading:** Managing FAISS indices across multiple modalities required optimized storage and retrieval logic for larger dataset.
- **Latency:** Real-time query execution introduced latency challenges, especially with multimodal embeddings.
- **Model Limitations:** CLIP's pretrained embeddings might not align perfectly with niche product descriptions, impacting retrieval precision.
- **Prompt :** GPT4 does not provide correct information on retrieved item.



Solutions

- **Fallback Mechanisms:** Introduced zero embeddings for invalid images to maintain consistency in index sizes.
- **Batch Processing:** Reduced memory overhead during embedding computation by processing data in manageable batches.
- **Pre-caching:** Cached indices and datasets using `st.cache_resource` in Streamlit to minimize runtime delays.
- **Hybrid Retrieval:** Designed a multimodal index to balance the strengths of text and image embeddings.
- **Enhanced Evaluation:** Developed a robust evaluation pipeline to measure accuracy for fine-tuning retrieval parameters.
- **Context Generation :** Used each item's name, description, category, price, and image URL as GTP4's context generation.

10. Future Work

- **Fine-tuning CLIP:** Train the CLIP model on domain-specific data to improve alignment for niche product descriptions.
- **FLAVA:** Try FLAVA model to generating embeddings and compare results with CLIP.
- **Index Optimization:** Experiment with FAISS's HNSW (Hierarchical Navigable Small World) for faster, scalable ANN searches.
- **Dynamic Index Updates:** Implement dynamic updates to FAISS indices for new product additions without requiring full index rebuilds.
- **Visual Enhancements:** Integrate a carousel or grid display for retrieved product images in the Streamlit UI.
- **Multilingual Support:** Extend the system to support queries in multiple languages, leveraging translation models like MarianMT.
- **Advanced Recommendations:** Explore techniques like cross-attention or transformer-based recommender systems to provide personalized results.
- **Cloud Deployment:** Host the entire application on cloud platforms (e.g., AWS, GCP) to support larger datasets and user bases.