Fashion Search AI

Prepared by: **Huy Vo Duc**

Course 6 - GenAI Elective – Module 11

upGrad and IIITB Machine Learning & AI Program – November 2024

Contents

1 Introduction

The purpose of this project is to design and implement a **Generative Semantic Search system** using a real-world fashion product dataset [Myntra dataset](#) on Kaggle. The system assists users in finding relevant clothing items based on **natural language queries**, combining:

- Semantic search (Retrieval)
- Intelligent ranking (Reranking)
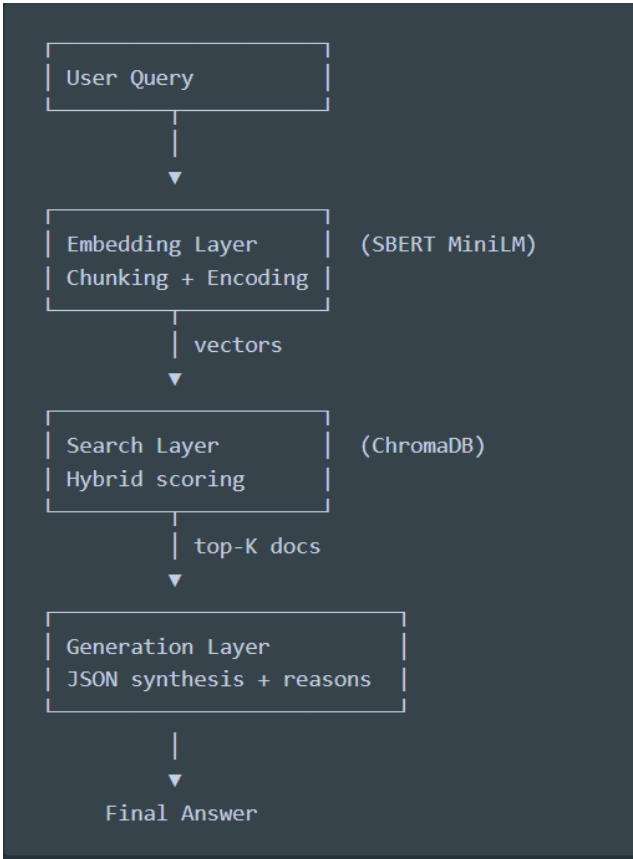- Justified recommendations (Generation)

---

2 Objectives

- Improve search relevance using semantic embeddings
- Provide **transparent explanations** for recommendation decisions
- Compare results between **Search Layer** and **Generation Layer**
- Experiment with chunking approaches, embedding models, and reranking strategies

---

3 Architecture Overview

The system contains **3 major components**:

Layer	Responsibility	Key Technology
Embedding Layer	Convert product text into vector representations	SentenceTransformers MiniLM
Search Layer	Retrieve and rank relevant products	ChromaDB + hybrid scoring
Generation Layer	Synthesize final structured answer with reasoning	OpenAI GPT and fallback Rule-based offline mode

```
┌─────────────────────┐
│ User Query          │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│ Embedding Layer     │   (SBERT MiniLM)
│ Chunking + Encoding │
└─────────────────────┘
          │ vectors
          ▼
┌─────────────────────┐
│ Search Layer        │   (ChromaDB)
│ Hybrid scoring      │
└─────────────────────┘
          │ top-K docs
          ▼
┌─────────────────────┐
│ Generation Layer    │
│ JSON synthesis + reasons │
└─────────────────────┘
          │
          ▼
     Final Answer
```

Architecture Diagram

4 Dataset Summary

- Source: **Myntra Fashion Dataset** (Kaggle)

- Type: Product metadata and descriptions

- #Rows: ~14,000 usable records

- Product images including

- Price currency: INR (Indian Rupees)

---

5 Design & Implementation

5.1 Embedding Layer

- Product-level chunk strategy (one chunk per item)

- Embedding model: "sentence-transformers/all-MiniLM-L6-v2"

- Stored in **persistent** ChromaDB
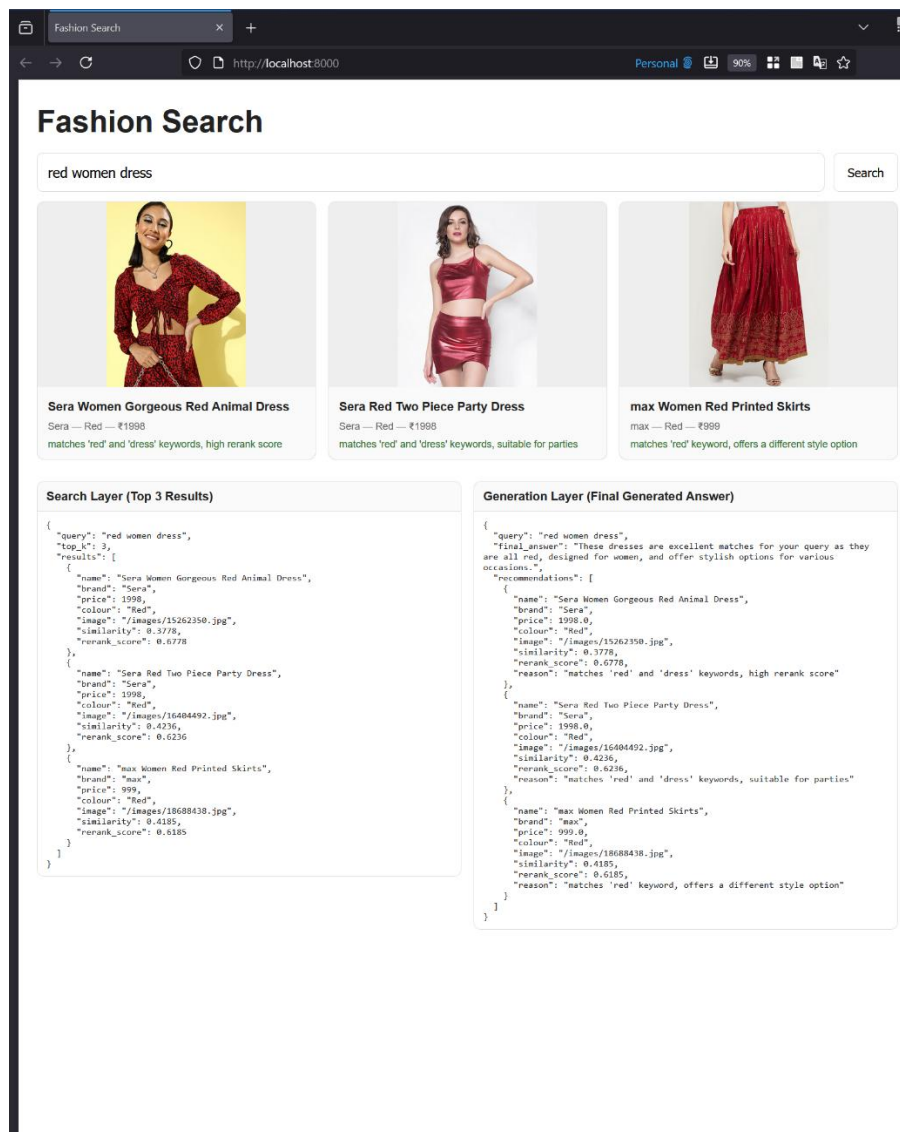
5.2 Search Layer

- Cosine similarity search

- Hybrid reranking with **keyword boosting**:

  - Color (e.g., "red", "blue")

  - Clothing type (dress, shoe, shirt)

  - Gender (men/women)

- Returns metadata including:

  - "similarity" and "rerank_score"

5.3 Generation Layer

- Produces "final_answer" summary

- Top-3 curated "recommendations"

- "reason" explains matching logic

- Optional LLM enhancement — "gpt-4o-mini"

- Automatic fallback: rule-based deterministic output – Offline mode possible

6 Features

- Fast search bar

- Spinner + skeleton loading

- Vertical product cards with full images

- Product details and recommendation reasons

- **JSON comparison panels (left: Search layer, right: Generation layer). This is for academy purposes which can be easy to differentiate**

Simple UI

7 Screenshot Of Output

The top 3 answers retrieved from the search layer, and the 3 final answers generated from the generation layer

Query	Search Layer	Generation Layer
"red women dress"	**Search Layer (Top 3 Results)** ```json{ "query": "red women dress", "top_k": 3, "results": [{ "name": "Sera Women Gorgeous Red Animal Dress", "brand": "Sera", "price": 1998, "colour": "Red", "image": "/images/15262350.jpg", "similarity": 0.3778, "rerank_score": 0.6778 }, { "name": "Sera Red Two Piece Party Dress", "brand": "Sera", "price": 1998, "colour": "Red", "image": "/images/16404492.jpg", "similarity": 0.4236, "rerank_score": 0.6236 }, { "name": "max Women Red Printed Skirts", "brand": "max", "price": 999, "colour": "Red", "image": "/images/18688438.jpg", "similarity": 0.4185, "rerank_score": 0.6185 }] }```	**Generation Layer (Final Generated Answer)** ```json{ "query": "red women dress", "final_answer": "These dresses are excellent matches for your query as they are all red, designed for women, and offer stylish options for various occasions.", "recommendations": [{ "name": "Sera Women Gorgeous Red Animal Dress", "brand": "Sera", "price": 1998.0, "colour": "Red", "image": "/images/15262350.jpg", "similarity": 0.3778, "rerank_score": 0.6778, "reason": "matches 'red' and 'dress' keywords, high rerank score" }, { "name": "Sera Red Two Piece Party Dress", "brand": "Sera", "price": 1998.0, "colour": "Red", "image": "/images/16404492.jpg", "similarity": 0.4236, "rerank_score": 0.6236, "reason": "matches 'red' and 'dress' keywords, suitable for parties" }, { "name": "max Women Red Printed Skirts", "brand": "max", "price": 999.0, "colour": "Red", "image": "/images/18688438.jpg", "similarity": 0.4185, "rerank_score": 0.6185, "reason": "matches 'red' keyword, offers a different style option" }] }```
"blue denim men shirt"	**Search Layer (Top 3 Results)** ```json{ "query": "blue denim shirt", "top_k": 3, "results": [{ "name": "URBANIC Women Blue Cotton Light Fade Colourblocked Jeans", "brand": "URBANIC", "price": 2090, "colour": "Blue", "image": "/images/16262928.jpg", "similarity": 0.6689, "rerank_score": 0.7689 }, { "name": "I Saw It First Women Blue Highly Distressed Jeans", "brand": "I Saw It First", "price": 4199, "colour": "Blue", "image": "/images/17403510.jpg", "similarity": 0.6611, "rerank_score": 0.7611 }, { "name": "URBANIC Women Navy Blue Cotton Relaxed Fit Mildly Distressed Jeans", "brand": "URBANIC", "price": 2190, "colour": "Navy Blue", "image": "/images/16262824.jpg", "similarity": 0.6516, "rerank_score": 0.7516 }] }```	**Generation Layer (Final Generated Answer)** ```json{ "query": "blue denim shirt", "final_answer": "The recommendations focus on blue denim items that complement the style of a blue denim shirt, featuring various shades and styles of jeans that can create a cohesive look.", "recommendations": [{ "name": "URBANIC Women Blue Cotton Light Fade Colourblocked Jeans", "brand": "URBANIC", "price": 2090.0, "colour": "Blue", "image": "/images/16262928.jpg", "similarity": 0.6689, "rerank_score": 0.7689, "reason": "Matched keywords 'blue' and 'denim', high similarity score." }, { "name": "I Saw It First Women Blue Highly Distressed Jeans", "brand": "I Saw It First", "price": 4199.0, "colour": "Blue", "image": "/images/17403510.jpg", "similarity": 0.6611, "rerank_score": 0.7611, "reason": "Matched keywords 'blue' and 'denim', high similarity score." }, { "name": "URBANIC Women Navy Blue Cotton Relaxed Fit Mildly Distressed Jeans", "brand": "URBANIC", "price": 2190.0, "colour": "Navy Blue", "image": "/images/16262824.jpg", "similarity": 0.6516, "rerank_score": 0.7516, "reason": "Matched keywords 'blue' and 'denim', relevant style." }] }```

Query	Search Layer	Generation Layer

Navi Jean less than price 2000	**Search Layer (Top 3 Results)**
	
	``json
	{
	  "query": "Navi Jean less than pirce 2000",
	  "top_k": 3,
	  "results": [
	    {
	      "name": "Ginger by Lifestyle Women Yellow Printed Sweatshirt",
	      "brand": "Ginger by Lifestyle",
	      "price": 999,
	      "colour": "Yellow",
	      "image": "/images/16091132.jpg",
	      "similarity": 0.3345,
	      "rerank_score": 0.3345
	    },
	    {
	      "name": "Ginger by Lifestyle Women Blue Printed Sweatshirt",
	      "brand": "Ginger by Lifestyle",
	      "price": 699,
	      "colour": "Blue",
	      "image": "/images/19189662.jpg",
	      "similarity": 0.3336,
	      "rerank_score": 0.3336
	    },
	    {
	      "name": "RAREISM Women Black Hooded Sweatshirt",
	      "brand": "RAREISM",
	      "price": 3299,
	      "colour": "Black",
	      "image": "/images/16533910.jpg",
	      "similarity": 0.3309,
	      "rerank_score": 0.3309
	    }
	  ]
	}
	``
		Generation Layer (Final Generated Answer)
		
		``json
		{
		  "query": "Navi Jean less than price 2000",
		  "final_answer": "While there are no jeans specifically matching your query, I recommend considering these affordable sweatshirts that offer style and comfort within your budget.",
		  "recommendations": [
		    {
		      "name": "Ginger by Lifestyle Women Yellow Printed Sweatshirt",
		      "brand": "Ginger by Lifestyle",
		      "price": 999.0,
		      "colour": "Yellow",
		      "image": "/images/16091132.jpg",
		      "similarity": 0.3345,
		      "rerank_score": 0.3345,
		      "reason": "Affordable price, good brand, and stylish design."
		    },
		    {
		      "name": "Ginger by Lifestyle Women Blue Printed Sweatshirt",
		      "brand": "Ginger by Lifestyle",
		      "price": 699.0,
		      "colour": "Blue",
		      "image": "/images/19189662.jpg",
		      "similarity": 0.3336,
		      "rerank_score": 0.3336,
		      "reason": "Affordable price, good brand, and stylish design."
		    }
		  ]
		}
		``

8 Evaluation & Sample Outputs

Layer	What it demonstrates
Search Layer	Pure retrieval + hybrid scoring (evidence)
Generation Layer	Explanation, reasoning, filtering & final curation

Improvements observed:

- More relevant product type filtering
- Better color/gender consistency
- Brand diversity in top-3 results
- Justification improves user trust

9 Challenges Faced

Issue	Resolution
Cross-encoder runtime errors	Replaced with custom hybrid scoring
ChromaDB API changes (ids removed)	Updated search interface
Torch "meta tensor" issue	Removed heavy CE model

10 Lessons Learned

- RAG systems must include robust fallback behavior

- Retrieval transparency improves trust and debugging

- Hybrid scoring is lightweight and effective vs. CrossEncoder

11 Conclusion

This project successfully implements a practical RAG-based generative search system that:

- Retrieves relevant fashion items
- Ranks them intelligently
- Explains its decisions in human-friendly language
- Works offline or enhanced by LLMs when available

12 References

- Dataset: Myntra Fashion Dataset v2 (Kaggle)

- SentenceTransformers (Reimers & Gurevych, 2019)

- ChromaDB (Vector Search System)

- FastAPI Documentation