# Fashion AI Chatbot – Project Report

## 1. Introduction

This project builds an AI-powered fashion search system using LangChain to enable intelligent, natural language-based product discovery. By leveraging Retrieval-Augmented Generation (RAG), vector search, and LLM-based query processing, the system searches through a vast collection of fashion product descriptions and recommends the most relevant choices.

The Myntra dataset from Kaggle serves as the primary data source.

## 2. System Design

The system consists of multiple layers that work together to provide intelligent fashion search capabilities:  
• Data Collection & Preprocessing  
• Text Embedding & Vectorization (LangChain + FAISS)  
• Retrieval-Augmented Generation (RAG) for Fashion Search  
• LLM-based Query Processing with LangChain  
• Conversational Agent for Interactive Search

## 3. Implementation

## Data Collection & Preprocessing

* + Load the Myntra dataset (CSV format) using Pandas.
  + Clean missing values and normalize text fields (prod id. price, name, descriptions)

## Text Embedding & Vectorization

* + Convert product descriptions into embeddings using OpenAIEmbeddings.
  + Store embeddings in FAISS for fast retrieval.

1. **Retrieval-Augmented Generation (RAG)**
   * Implement semantic search using LangChain’s VectorStoreRetriever.
   * Retrieve similar product descriptions based on user queries.
2. **LLM-based Query Processing**
   * Use LangChain’s LLMChain to refine and interpret user queries.
   * Convert user-friendly queries into structured search parameters.
3. **Conversational AI Agent**
   * Integrate LangChain’s ConversationalRetrievalChain to allow follow-up queries.
   * Maintain context: Users can refine searches dynamically.

## 4. System Flowchart

The following diagram represents the overall system architecture:

LLM Processing

Vector Search (FAISS)

Response Generation

(Fashion Recommendations)

## 5. Challenges & Lessons Learned

• Managing large-scale embeddings required optimizing FAISS retrieval.

• Ensuring LLM-generated responses aligned with user expectations.

• Fine-tuning query processing improved accuracy in recommendations

## 6.Instructions for running the notebook

* Use openai key for running
* Use Myntra dataset from Kaggle as the dataset.(provided the link in the notebook)