Product Search Project Documentation

# 1. Project Overview

This project allows searching products using:  
- Text search (SentenceTransformer all-MiniLM-L6-v2)  
- Image search (CLIP model clip-ViT-B-32)  
- Products embedded & stored in ChromaDB  
- Backend APIs with FastAPI for Insert, Update, Delete, Search

# 2. Project Structure

AI\_PRODUCT\_SEARCH\_MAIN/  
 - backend.py # FastAPI backend (CRUD + Search APIs)  
 - embed\_to\_chroma.py # Embedding script  
 - chroma\_db/ # ChromaDB storage

# 3. Setup Instructions

1. Install Python 3.9+  
2. Create virtual environment:  
 python -m venv venv  
 source venv/bin/activate (Linux/Mac)  
 venv\Scripts\activate (Windows)  
  
3. Install dependencies (single command):  
 pip install fastapi uvicorn chromadb psycopg2 sentence-transformers pillow requests aiohttp

Or

pip install -r requirements.txt

**4. Enable GPU Support (Optional for Faster Embedding):**

1. Ensure an NVIDIA GPU is installed and update the NVIDIA driver from [NVIDIA’s Download page](http://www.nvidia.com/Download/index.aspx) to the latest version (e.g., 535 or higher).

2. Uninstall existing PyTorch: pip uninstall torch torchvision

3. Install PyTorch with CUDA support (e.g., CUDA 11.8): pip install torch torchvision --index-url https://download.pytorch.org/whl/cu118

4. Verify GPU availability: python -c "import torch; print(torch.cuda.is\_available()); print(torch.cuda.get\_device\_name(0))" //should output True and GPU name//.

5. Note: If no GPU or CUDA setup is available, the script defaults to CPU, which is slower.

# 5. Data Embedding

Run: python embed\_to\_chroma.py  
- Fetch products from PostgreSQL  
- Embed text & images  
- Store embeddings in ChromaDB collections

# 6. Backend API (FastAPI)

Run server:  
 uvicorn backend:app --reload  
  
Docs available at:  
 http://127.0.0.1:8000/docs  
 http://127.0.0.1:8000/redoc

# 7. Available APIs

Root: POST /  
Fetch products: POST /products (with offset & limit)

Text search: POST /search (with query payload)

Image search: POST /image-search (upload file)

Insert: POST /insert

Update: POST /update

Delete: POST /delete

## Payload Examples

Insert Payload Example:

{  
 "id": "12345",  
 "oem\_id": "OEM-111",  
 "name": "Test Product",  
 "description": "This is a test insert",  
 "images": "https://example.com/img.jpg",  
 "specifications": "Spec A, Spec B"  
}

Update Payload Example:

{  
 "id": "12345",  
 "name": "Updated Product",  
 "description": "Updated desc",  
 "images": "https://example.com/new.jpg",  
 "specifications": "Spec X, Spec Y"  
}

Delete Payload Example:

{  
 "id": "12345"  
}

# 8. Testing APIs

- Use Postman or curl  
- Insert/Update/Delete: Method POST → Body → raw → JSON  
- Image Search: Method POST → Body → form-data → file

# 9. Project Flow

1. PostgreSQL → embed\_to\_chroma.py → ChromaDB  
2. backend.py → FastAPI → CRUD + search  
3. Frontend (optional) → API calls → UI results