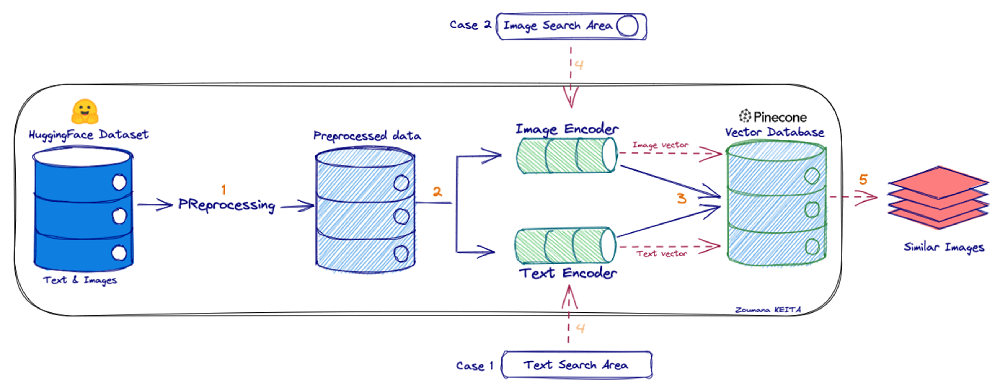
Our Approach – Retrieval Augmented Generation

* For visual search, we’ve encoded query images using the trained encoder and then find similar encoded representations from our dataset.
* OpenAI's CLIP (Contrastive Language-Image Pretraining) model is used for embedding images. CLIP is suitable for text image pair-based tasks, and we’ve made use of text-image and image-image search. The model uses a ViT-B/32 Vision Transformer architecture as an image encoder and uses a masked self-attention Transformer as a text encoder.
* Each of these embeddings are then saved to Pinecone Vector Database which is hosted in free tier of GCP for semantic search against query. We make access request to this cloud hosted db via Pinecone Python Client package. Pinecone Vector DB consists of 512-dimension space to fit vectors. We first upsert vectors and then run query to match vectors using cosine similarity method. Thereby fetching similarity scores.
* For product discovery, we generate new product images using the trained generator and display them to users based on their preferences.

Reference:

1. Technique - <https://www.pinecone.io/learn/clip-image-search/>
2. Model - <https://huggingface.co/openai/clip-vit-base-patch32>
3. Dataset - <https://huggingface.co/datasets/ashraq/fashion-product-images-small>
4. Pinecone API - <https://docs.pinecone.io/docs/image-similarity-search>
5. Open API- https://openai.com/research/clip

Workflow: