ODSC WEST
LIGHTNING TALKS

SAN FRANCISCO
29-31 OCTOBER

SHUBHAM GOEL

Senior Machine Learning Scientist
ZEFR

LABELLING SPARSE DATA AT SCALE USING SEMANTIC SEARCH

Problem

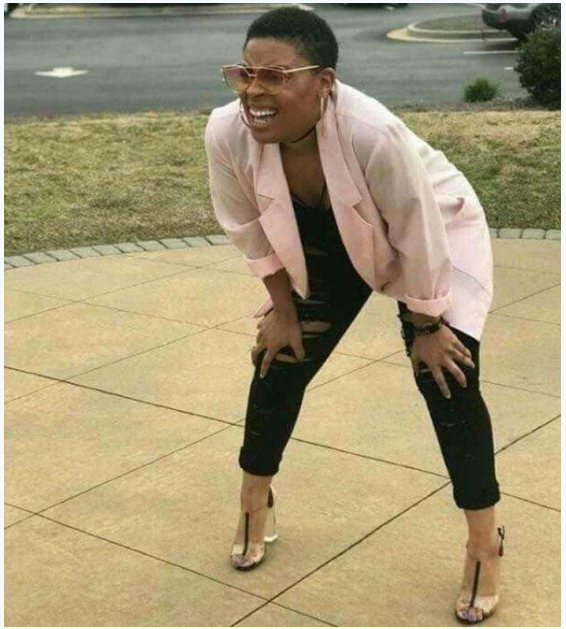
- Brand Safety across platforms - **YouTube, TikTok, Meta, Snap**

- Brands such as Adidas, Disney, etc. looking to advertise on these platforms come to us to ensure **Brand Suitability** and **Safety** across several risk categories
    - Profanity
    - Hate speech
    - Adult
    - Drugs, Alcohol and Tobacco
    - Graphic
    - etc.

- Need robust models to classify content using multiple features -
    - Caption
    - Hashtags
    - Video
    - Audio transcript
    - User/Content metadata such as likes, impressions, user handle, etc.

"81% of respondents would stop purchasing a product they regularly buy if they discovered the brand's ads had appear next to racist content or hate speech"

Data, Data, Everywhere!

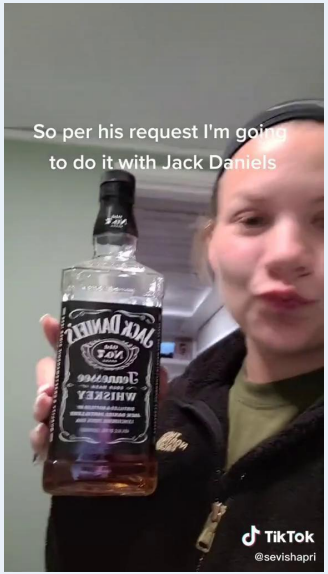


Visual depiction of me trying to find anything in that data

What if you’re looking for a video which has

- a guy holding a beer can
- someone talking about doing drugs
- people texting about sexually explicit content
- missiles being fired in the background
- two people dressed up in deadpool cosplay suits fighting with swords
- ...(let your imagination run wild)

Or maybe a video similar to



Semantic Search be like

Text search

two people dressed up in deadpool cosplay suits fighting with swords

Download IMAGE

Similarity: 0.4618

Platform ID: 7043071623586467074/111.jpg Video Link

Download IMAGE

Similarity: 0.4604

Platform ID: 7019963413400964353/416.jpg Video Link

Download IMAGE

Similarity: 0.4599

Platform ID: 7025645755851410693/293.jpg Video Link

Image search

Download IMAGE

Similarity: 0.6052

Platform ID: 6960934672473869573/414.jpg Video Link

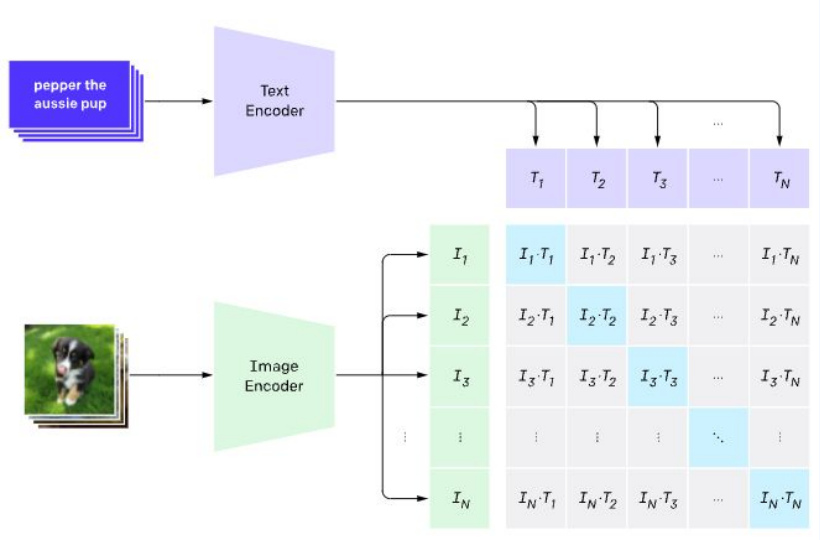
Download IMAGE

Similarity: 0.6021

Platform ID: 7042055333207969029/148.jpg Video Link

OpenAI's CLIP for extracting embeddings

- Pre-trained on 400M Image-Text pairs with Contrastive loss
- Objective:
  - *Minimize* distance between image/text embeddings from the same example, AND
  - Maximize distance between image/text embeddings from different examples
- Modular and relatively light-weight
- Multiple variants of CLIP available
  - Using the LAION-2B L/14 model currently

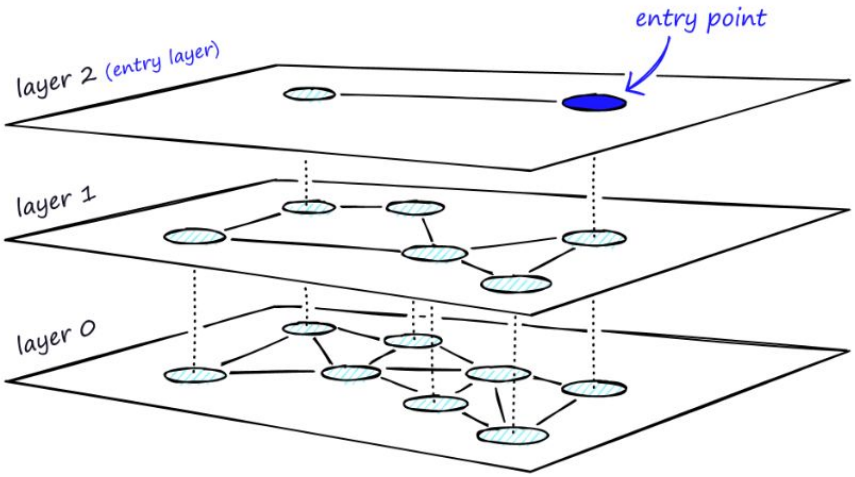


How to index?

- Index embeddings extracted for all frames (TikTok, YouTube, Meta)
- Have fast response times (<100ms)
- Probably okay to do an approximate search
  - Top 8 out of 10 best results works, especially as the index size scales

Approximate Nearest Neighbors (ANN) to the rescue

Hierarchical Navigable Small Worlds (HNSW) Graph

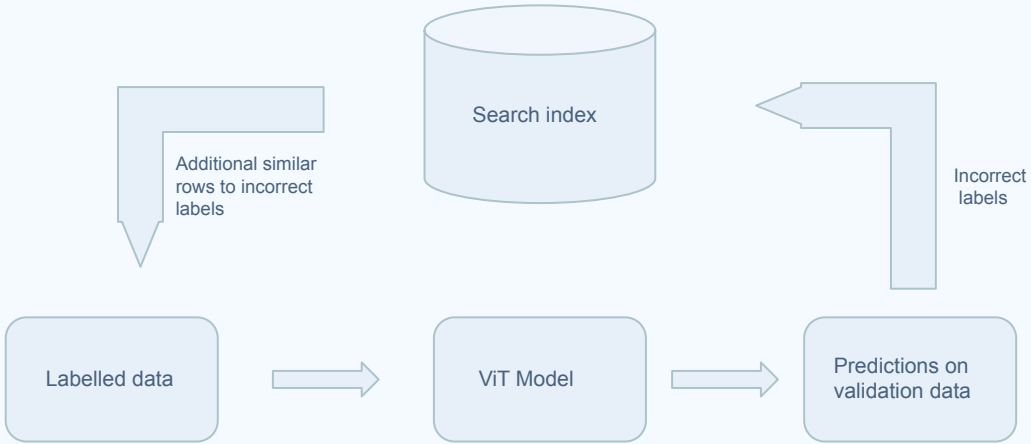


Layered graph of HNSW, the top layer is our entry point and contains only the *longest links*, as we move down the layers, the link lengths become shorter and more numerous.

Qdrant

- Open-source, scalable and efficient
- Allows addition, deletion, querying together at serving time
  - Optimized for real-time serving
- Supports a variant of HNSW; modified to support real-time updates
- Allows complex filtering queries with metadata, combining vector search with query conditions
  - ***"find images similar to `Image A` which contain the text 'fyp' in hashtags and is >30 secs"***
- Has Custom Ranking function support, to further fine-tune results based on specific requirements

How is it being used?



Future Goals

- Large scale fine tuning of CLIP ViT model
  - Tiktok/Meta video/caption pairs
  - Filter the best matching frame to the caption using CLIP Similarity

- 200M image-caption pairs (growing daily)
  - Filter additional ones out based on low similarity score.
  - Plan to scale to >1B vectors very soon

- Add additional metadata to enable even more complex filters

- Finetune and re-index!


THANK YOU FOR YOUR ATTENTION
PLEASE CLAP AND DON'T ASK ANY DIFFICULT QUESTIONS
memegenerator.net