**Springboard: Data Science Career Track Program**
**Project 3 - Detecting AI-Generated Images with Classification, Similarity analysis, and Clustering**
**Proposal By Priyanka Panhalkar**
**[March], [2025]**

## Business Problem

With the rise of Generative AI (GenAI), distinguishing between real and AI-generated images is challenging. This leads to issues like misinformation, fraud, and deep fake misuse. We aim to develop an AI model that automatically detects whether an image is real or AI-generated, improving trust, security, and transparency across industries.

## Intended Stakeholders

- **Social Media Platforms (Facebook, Instagram, TikTok, Twitter/X):** Flag AI-generated content to prevent misinformation.
- **News & Media Companies:** Ensure published images are authentic.
- **E-Commerce Platforms (Amazon, eBay, Etsy)**: Prevent AI-generated product misrepresentation.
- **Cybersecurity Teams:** Detects deep fake scams and fraudulent images.

## Dataset Collection

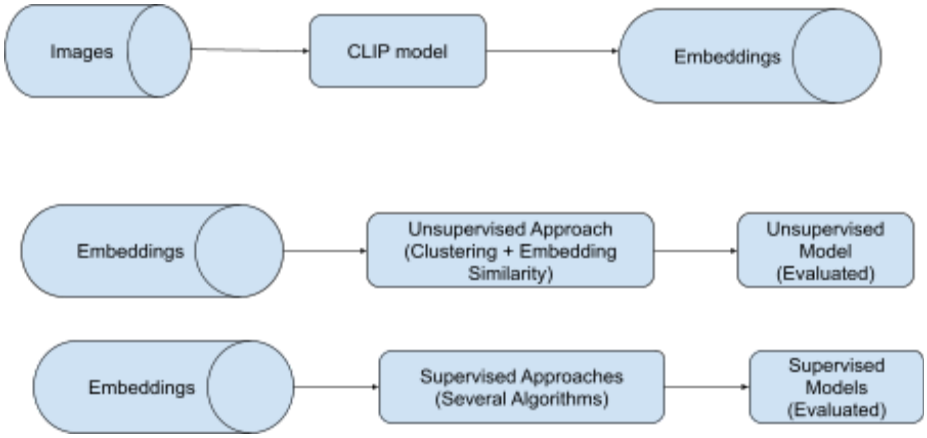
I came up with 50 different diverse search queries like Golden Gate Bridge or Beaches etc. and did searches on Pinterest for real images and AI generated images. I created a structured pipeline to collect real and AI-generated images from Pinterest by extracting image URLs using regex "*https:\/\/i\.pinimg\.com\/[^"'\s]+?\.jpg*" from search page HTML. These URLs were stored in a dataset, loaded into a pandas DataFrame, and downloaded to Google Drive via Colab.

**Failed Attempts** - As for AI generation, I also attempted to use Stable Diffusion on Colab but found the image quality unsatisfactory.

## Data Science Approach

### 1. Preprocessing & Feature Extraction:

Resize, normalize, and augment images to boost model robustness. We will use the CLIP Model from openAI to convert images into 512-dimensional embeddings




Once we create the embeddings we will using two different approaches as below:

2. **Unsupervised Model**

   We will be performing an unsupervised approach as below

   - Cosine Similarity matrix clustering Calculate cosine similarity between embeddings (e.g., a cosine value of 0.98 indicates high similarity).
   - Apply an unsupervised Clustering Algorithm eg K-Means to group similar images.
   - Use **t-SNE** to reduce the high-dimensional embeddings into 2D or 3D.

3. **Supervised Model**

   We will be performing an supervised approach as below

   - **Classification:** Treat the embeddings as feature inputs for classifiers (e.g., Logistic Regression, SVM, Random Forest, or deep models like EfficientNet/ViTs). Evaluate using a confusion matrix, classification report, and precision-recall curves to determine whether an image is AI-generated or real.

All the Models built will be using the same training and test Set. Also, because the images have been labeled we will be able to compare Precision, Recall, Confusion Matrix and Classification Report , Precision Recall Curve.

Important Links:

1. https://github.com/openai/CLIP
2. https://github.com/openai/CLIP/blob/main/CLIP.png