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.

gitgitDataset Collection

To train the model, we need **two types of images**:

1. **Real Images** – From trusted sources like ImageNet, CelebA, Pinterest
2. **AI-Generated Images** – From GenAI models like **Stable Diffusion, Deepfake, GANs**.

Web Scraping for More Data

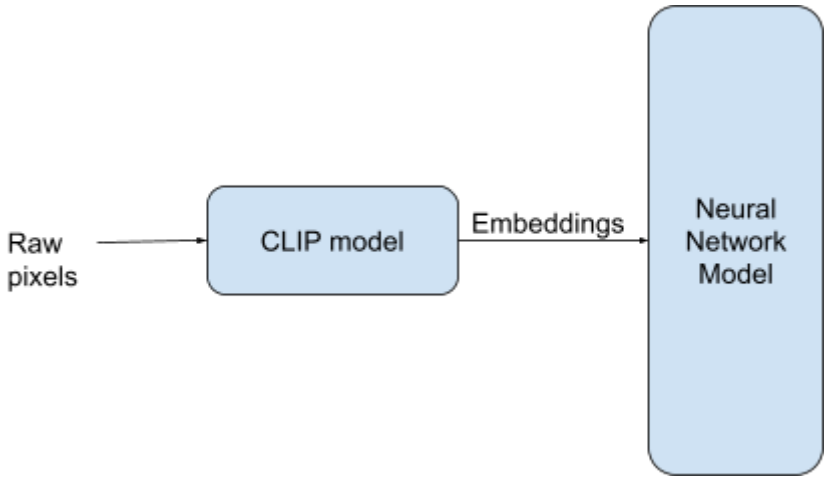
1. **Find AI-generated image sources**:

   - **Hugging Face Spaces** (AI image demos)
   - **Lexica.art** (Stable Diffusion search)
   - **Pinterest**

Data Science Approach

Preprocessing & Feature Extraction:

Resize, normalize, and augment images to boost model robustness. Use CLIP to convert images into 512-dimensional embeddings.



Using Embeddings for Classification & Similarity Analysis:

1. **Similarity Measures:** Calculate cosine similarity between embeddings (e.g., a cosine value of 0.98 indicates high similarity). Visualize the embedding space using PCA or t-SNE to reveal clustering.
2. **Clustering:** Apply clustering algorithms like K-Means or DBSCAN on the embeddings using Euclidean distance. This groups AI-generated images by similarity to real ones, potentially ranking images by realism.
3. **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.

This concise approach integrates preprocessing, feature extraction with CLIP, similarity analysis, clustering, and classification for robust image analysis.

Important Links:

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