Capstone Three- Project Ideas

Idea 1: Classify the image if its GenAI or Real

In an era where artificial intelligence (AI) is revolutionizing various industries, the ability to discern between real and AI-generated content has become increasingly crucial. The prevalence of Generative AI (GenAI) has brought about innovative advancements, particularly in creating realistic images that are often indistinguishable from those captured by conventional means. While this technology has opened up new avenues for creativity and automation, it has also introduced challenges in identifying and verifying the authenticity of such images.

The primary objective of this project is to develop a robust system capable of determining whether an image has been generated by GenAI or not. This capability is essential in several domains, including digital forensics, media integrity, and content authentication.

A) Business Problem

With the rapid advancements in **Generative AI (GenAI)**, it is becoming increasingly difficult to distinguish between **AI-generated images** and **real images**. This raises significant concerns in various industries, including **media, cybersecurity, e-commerce, and digital content verification**.

The **main business problem** is:

- How can we **automatically classify** whether an image is AI-generated or real?
- How can organizations **prevent misinformation, deep fake misuse, and fraud** by leveraging AI-based detection methods?

By building a **GenAI vs. Real Image Classification Model**, we can **improve trust, security, and transparency** in digital content.

B) Intended Stakeholders

- **Social Media Platforms** (Facebook, Instagram, TikTok) – Prevent AI-generated misinformation and fake content.

- **News & Media Companies** – Ensure published images are real and not AI-manipulated.
- **E-Commerce (Amazon, eBay, Etsy)** – Stop sellers from using AI-generated images that misrepresent products.
- **Cybersecurity Teams** – Detect deepfake scams and fraud in online identity verification.

C) Where is the dataset available from?

To train the model, we need **two types of images**:
1. **Real Images** – From trusted sources like ImageNet, COCO, CelebA.
2. **AI-Generated Images** – From GenAI models like **DALL·E, MidJourney, Stable Diffusion, Deepfake, GANs**.

Sources for AI Images:

Public datasets:

- This Person Does Not Exist (AI-generated faces)
- Deepfake Detection Challenge Dataset
- Generated vs. Real Image datasets from Kaggle

Web Scraping for More Data

1. **Find AI-generated image sources**:
    - **Hugging Face Spaces** (AI image demos)
    - **Lexica.art** (Stable Diffusion search)
    - **Reddit** (r/StableDiffusion, r/AIArt)
2. **Use Web Scraping Tools**:
    - **BeautifulSoup** (parses HTML)
    - **Scrapy** (for large-scale scraping)
    - **Selenium** (for dynamic pages)
3. **Filter AI-Generated Images**:
    - Look for **watermarks** from AI models.
    - Analyze **metadata** (EXIF data, filenames with "AI-generated").
    - Use **Reverse Image Search (Google Lens, Bing)** to find similar AI-generated images.

D) Data Science Approach

We will rely on 3rd party zero shot learning models like CLIP from openAI to generate embeddings out of raw pixels and use those embeddings in the model training and evaluation. CLIP

Model Architecture

We will use the Deep neural network classification model. We will use the pytorch library to create this neural network.

Evaluation

Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.

Cross-validation & testing on unseen datasets to assess real-world performance.