ML HACKATHON

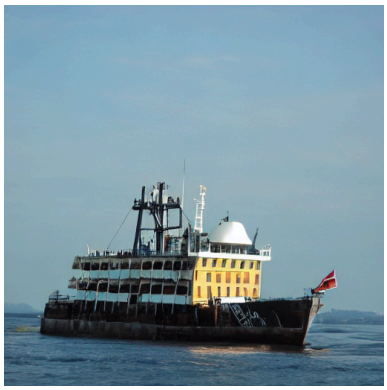
Introduction:

The rapid advancement of artificial intelligence has transformed digital content creation. Cutting-edge generative models like Adobe Firefly, Stable Diffusion, and Midjourney now enable the production of highly realistic synthetic media. These diffusion-based models unlock new creative opportunities and offer businesses efficient solutions for automating content generation. However, the rise of synthetic media also brings significant challenges regarding content authenticity and trustworthiness.

The task of detecting AI-generated images becomes tough when adversarial perturbations are added to the image. Handling these perturbations is very challenging, and it is important to differentiate between a real and AI-generated image.

Problem Statement:

Build a model that accurately identifies AI-generated images, focusing on those generated by diffusion and other advanced techniques.
Task: Detection of AI-Generated Images.
Goal: Classify images as either AI-generated or real. A data sample is provided below.
The dataset consists of labelled examples of real and fake images
REAL: containing real images
FAKE: containing synthetically generated images



This is a Fake image.

This is a Real image.

The results will be evaluated on custom test datasets, which will be shared 2 hours prior to final submission.
Note: The final test dataset images will be scaled down to 32x32, i.e., the same dimension as images in the CIFAKE dataset.

Dataset:

The link to the dataset is Dataset.

The dataset consists of 50k FAKE images and 50k REAL images for training and 10k REAL and 10k FAKE images for testing.

Evaluation:

The evaluation will be done on two test datasets.
Test_dataset_1: Contains both real and fake images without adversarial perturbations.
Test_dataset_2: Contains both real and fake images with adversarial perturbations.

Evaluation metrics:
Classification Accuracy, Precision, Recall, and F1 Score.

Note: Results on Test_dataset_1 will contribute to 40% and Test_dataset_2 will contribute to 60% of the final evaluation.

Deliverables:

- All the code files.
- Weights of models.
- Readme file consisting of all necessary details.
- Two submission CSV files (Test_1_results.csv and Test_2_results.csv) with the structure shown below.

Image name	Prediction
1	Real
2	Real
3	Real
4	Real

Guidelines:

- Prediction should be “Real” or “Fake”.
- Submission will be done through a Google Form.
- All the deliverables should be submitted in a zip file with name format as TEAM_NAME_PVH_ML.zip
- Plagiarism will lead to disqualification.
- Mention all the details if using any pre-trained model.