

Data Collection and Preprocessing Phase

Date	29 November 2024
Team ID	739984
Project Title	AI-Generated vs. Real Image Classifier Using Deep Learning.
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The project focuses on building a classifier to distinguish between real and AI-generated images, using deep learning techniques such as CNN and transfer learning.
Data Collection Plan	The dataset will be sourced from Kaggle, containing labeled images of both AI-generated and real images, ensuring sufficient variety and size for model training.
Raw Data Sources Identified	Raw data will be collected from various Kaggle datasets, including "AI-Generated vs Real Images" and others that contain diverse images from GANs, AI platforms, and real-world photography.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Labeled dataset of real and AI-generated images for training image classifiers. The dataset contains various AI-generated image types and real-world photos.	https://www.kaggle.com/datasets/cashbowman/ai-generated-images-vs-real-images	Image	422MB	Public