

# AI-Generated Images vs. Real Images Classification

[Git-Hub Repository](#)

Sidiropoulos Michalis 2320

## Problem Statement

With the rise of artificial intelligence (AI) capabilities in generating realistic images, distinguishing between AI-generated images and real images has become increasingly challenging. This project aims to develop a deep learning model capable of accurately classifying images into two categories: AI-generated and real images.

## Motivation & Challenges

The proliferation of AI-generated content across various platforms raises concerns about misinformation and authenticity. Robust classification models are crucial for verifying the authenticity of images in applications such as content moderation, forensic analysis, and counterfeit detection. However, distinguishing between AI-generated and real images poses several challenges, including the rapid advancement of AI technology, the diversity of image styles and content, and the potential for subtle differences between AI-generated and real images.

## Approach

The proposed approach involves leveraging convolutional neural networks (CNNs), a powerful deep learning architecture widely used for image classification tasks. By training a CNN on a dataset containing both AI-generated and real images, the model will learn to differentiate between the two categories based on their distinct features and characteristics. Transfer learning techniques may be explored to enhance model performance, considering the limited availability of labeled data for AI-generated images.

## Baselines

As a baseline, traditional machine learning classifiers such as Support Vector Machines (SVM) and Random Forests will be implemented and compared against deep learning models. Additionally, a simpler CNN architecture like LeNet-5 will serve as baseline for evaluating the performance of more complex models.

## Dataset

The dataset for this project consists of a curated collection of AI-generated images and real images sourced from Kaggle. The dataset includes diverse image categories and encompasses a range of styles, resolutions, and content types to ensure model robustness and generalization. Specifically, the dataset is a captivating ensemble of images sourced from two distinct channels:

- Web scraping
- AI-generated content

The content covers many subjects however, special emphasis was placed on the following topics:

- people
- animals
- portraits
- scenery
- psychedelics

## Metrics

Performance evaluation will be conducted using standard classification metrics such as accuracy, precision, recall, and F1 score. Additionally, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores will provide insights into model discrimination capabilities and performance across different thresholds.