Project C9: KAGGLE- AI IMAGE DETECTOR

Team Members: Jakko Turro, Lauri Laud, Markus Tõnson

GitHub repo: https://github.com/JakkoT/IDS-AI-Image-detection

Task 2. Business Understanding

2.1 Identifying your business goals

2.1.1 Background

The rapid advancement of generative AI has led to an increase of AI-made artwork across all digital platforms. While this advances the creativity of some people, it also raises concerns regarding authenticity, copyright, transparency, and consumer trust. AI tools like Midjourney, DALL-E and Nano Banana have become very good at making fake images. This creates problems like deepfakes, fake news, and people claiming AI art is human-made. It is becoming hard for regular people to tell the difference.

2.1.2 Business goals

Our goal is to build a tool that can automatically check if an image is real or made by AI.

2.1.3 Business success criteria

The project is successful if our system can correctly flag AI images with about 75% accuracy.

2.2 Assessing the situation

2.2.1 Inventory of resources:

Data**:** We have a dataset with 40,000 images (20,000 Real, 20,000 AI).

Tools**:** We will use Python, PyTorch for the AI model, and Jupyter Notebooks for testing code.

Hardware**:** We will use our laptops with GPUs to train the model.

2.2.2 Requirements, assumptions, and constraints:

Requirement*:* Large and diverse datasets representing both human and AI art. The model must work with standard image files like JPG and PNG.

Assumption*:* We assume the "Real" images in our dataset are actually real and not edited.

Constraint: Rapid evolution in generative AI may cause the model to become outdated. We are limited by our computer power and the course deadline.

2.2.3 Risks and contingencies:

Risk: AI generators change fast. Our model might work on today's AI images but fail on next year's versions.

Risk: It is hard to explain why the model thinks an image is fake (Black Box problem).

2.2.4 Terminology:

True Positive - Correctly catching an AI image.

False Positive - Accusing a real image of being AI.

AI-generated/AI art **-** Artwork created using generative AI models.

Human-made/Real art **-** Artwork created by human artists.

2.2.5 Costs and benefits:

Costs: Time spent coding and training the model. (Maybe the electricity consumption also)

Benefits: So that we can differentiate between “AI” and “Real” images.

2.3 Defining your data-mining goals

2.3.1 Data-mining goals

We want to build a Convolutional Neural Network (CNN) using PyTorch. The model needs to find hidden patterns in the pixels that humans can't see to decide if an image is "Real" or "AI."

2.3.2 Data-mining success criteria

We aim for an accuracy of at least 75% on our test data. We also want a high F1-score to make sure we are balancing our predictions well.

Task 3. Data Understanding

3.1 Gathering data

3.1.1 Outline data requirements

We need a large set of images labeled "Real" and "AI." They need to be clear enough for the computer to analyze pixel patterns.

The dataset is from Kaggle and can be found here:
https://www.kaggle.com/datasets/mkevinrinaldi/my-sampled-art-dataset-40k

3.1.2 Verify data availability

Since we already have the dataset, it contains 40,000 images (20,000 human-made and 20,000 ai-generated) sorted into folders “Real” and "Fake". This is a quite substantial dataset size for training a deep learning model with good generalization.

3.1.3 Define selection criteria

We have selected the entire dataset for use. In Deep Learning, data volume is very important for feature extraction, and 40,000 images provide a strong baseline. Additionally, the dataset is perfectly balanced (50/50 split). We are not filtering by image subject because we want a general-purpose detector that performs well across various contexts, environments, and image styles.

3.2 Describing data

Volume: 40,000 image files.

Format: Standard RGB images (JPG/PNG).

Labels: The target variable is binary:

Real (1) - 20,000 images captured by cameras or created by human digital artists.

AI generated (0) - 20,000 images synthesized by generative models (likely Midjourney, Stable Diffusion or DALL-E).

Structure: Unstructured image data with varying art styles (modernism, expressionism, realism and others).

3.3 Exploring data

We will use Jupyter Notebooks to explore the files.

Visual Check: We will plot 10 random images from each folder to see what they look like. We want to see if there are obvious differences, like weird hands in AI images or oddly consistent stylistic patterns.

Size Analysis: We will check the height and width of the images. If they are all different sizes, we need to decide on a standard size (like 200x200 pixels) to resize them to before training.

Balance: We confirmed the split is exactly 50% Real and 50% AI, so we don't need to worry about biased data.

3.4 Verifying data quality

Corrupt Files: We will write a script to open every image and do some checks manually. If an image gives an error, we will delete it so it doesn't crash our training loop.

Duplicates: We will check if the same image appears twice. We definitely don't want the same image in both the "Training" and "Testing" sets, as that is cheating with the accuracy.

Label Check: We will manually look at a small batch to make sure the files in the "Real" folder are actually real and files in “Fake” folder are actually made by AI and not mislabeled.

Task 4. Planning your project

Project Plan

We will use this workflow for this project.

Task No.	Task	Description	Jakko	Lauri	Markus	Total
1	Project planning	Google Docs, GitHub	2	2	2	6
2	Data Prep & Cleaning	Jupyter. Check for bad files, resize images, and split into Train/Test sets.	5	5	5	15
3	Build Model	TensorFlow. Create the CNN layers (Convolution, Pooling, Dense).	6	6	6	18
4	Training	TensorFlow. Run the training loop. Adjust settings (learning rate, epochs) to get better results.	10	10	10	30
5	Testing & Graphs	Matplotlib. Test the model on new data. Create graphs to show Accuracy and Loss.	5	5	5	15
6	Final Report	Word. Write down our findings and present the results.	2	2	2	6
Total			30	30	30	90