# Detecting Images Generated by Neural Networks

**Team Members:** Alina Al Beaini, Amanda Pan, Cemile Kurkoglu, and Hasan Saad

**GitHub**: https://github.com/Alina-Beaini/AIvsReal

**Overview:** The recent advances in deep learning, neural networks, and the hardware to support it has provided fertile ground for creating fake images. This new technology, if left unchallenged, creates a risk in many areas including journalism and law enforcement. We tackle this problem by constructing two multi-classification models (single-channel and dual-channel) to discern between images which are real, that is, not generated by AI, and those which are generated by AI, and to determine which generative algorithm was used. Our model is trained on a publicly available dataset of ≈90000 images. This dataset contains real images as well as images generated by 13 different CNN-based generative algorithms.

**Stakeholders:** Social media companies, journalism organizations, law enforcement.

**Approach:** Our models are a single-channel and dual-channel Convolutional Neural Networks. In each model; we have ReLu for the activation function for output layer, Adam optimizer, categorical cross-entropy as the loss function. In the single-channel model, we have High Pass Filter using Gaussian blur. In the dual-channel one, one copy of the image has a High Pass Filter while the other copy has a Log-Scale and Normalized Discrete Cosine Transform.

To process our images, we use a python script which discards images below a certain dimension as well as grayscale images, and crops images to a fixed size. Furthermore, it saves the files into a new folder which contains the newly processed images, with filenames that include the necessary output material.

Since loading all the images into memory before training the neural network is not feasible, we create a custom Sequence element which reads the paths, and puts the images in memory on a "need-to-know" basis. In other words, on any given batch, only that batch is loaded into memory.

Since Keras does not natively support precision and recall metrics when more classes than two are involved, we write our own metrics.

Finally, we've written a web app that showcases our model.

**Future Iterations:** We plan to improve our model for detecting AI-generated images in terms of generalization to different methods of image generation and robustness to image resizing and compression.

Work on a more homogeneous dataset in terms of model/method that generated the images to improve precision and recall. We can use Co-occurrence matrices to build a neural network that identifies real versus fake images.