Missouri University of Science & Technology — Department of Computer Science

Spring 2024 — **CS 6406: Machine Learning for Computer Vision (Sec: 101/102)**

Homework 4: Trustworthy Vision

Instructor: *Sid Nadendla* — **Due:** *May 3, 2024*

Goals and Directions:

- The main goal of this assignment is to design and analyze trustworthy NNs using explainability tools available in Tensorflow.
- A template Jupyter notebook will be provided for each problem to develop your solution. Examples will be included in the same Jupyter notebook for reference.
- Although this assignment does not demand significant compute power, you may obtain any compute power needed from Foundry, Google Colab, or AWS SageMaker Studio Lab.

Problem 1 Testing with CAVs *13 points*

Import a pretrained CNN model of your choice (e.g. Resnet-50) from Keras library and perform transfer learning on CIFAR-10 dataset. Test a concept ***cloud*** in your model's ability to detect the class "*airplanes*". You may collect the cloud images from CIFAR-100 dataset. This enables you to understand if your model is truly learning the correct features in detecting an aircraft.

Problem 2 Fairness *12 points*

Import InceptionV3 pretrained model from Keras and perform transfer learning on UTKFace dataset (`https://susanqq.github.io/UTKFace/`) to predict *race* (White, Black, Asian, or Others) of each image. You need to download the dataset from the given link to your local machine. Evaluate the statistical parity, equalized odds and calibration of the model by evaluating the conditional probability of positive predictions across each race. Discuss your results and present your own fairness analysis for InceptionV3 pretrained model.