Real or Fake? Using Artificial Intelligence to Detect Filtered Images

Otten, Tamara (School: Mountain Vista Governor's School)

Misinformation is widespread on the internet, and one of the most pervasive ways this presents itself to adolescents is in the form of images manipulated on social media. Teens post pictures on Instagram or Snapchat pretending to look perfect, when in reality, the images have been heavily edited. These enhancements create unreasonable beauty expectations and lead to viewer insecurities and feelings of low self-esteem. In this research, we aim to address this issue by training a supervised two-way AI visual classifier with a large dataset of facial images. We introduce a dataset of filtered images, and train alongside their unaltered counterparts. Using this machine learning model, we were able to accurately classify these photos into their respective categories with over a 91% success rate. With further training and testing, this technique could likely be utilized to achieve higher performance on an even wider variety of filters. Currently, we are working to train additional models on several other filter types. These could be integrated into a pipeline to test for various filter combinations in order to reach the goal of a generalizable filtration detection system. Improved methods such as these could have profound impacts on the future and how we view the online world. Giving teens an easy way to determine if what they are looking at is fake can improve their mental health status and remind them that they should not compare themselves to these unrealistic standards. This provides a new approach to the ongoing battle to fight the spread of misinformation on the internet. These topics are of utmost importance, especially in a world where fabrication and deepfakes are becoming increasingly common.

Awards Won: