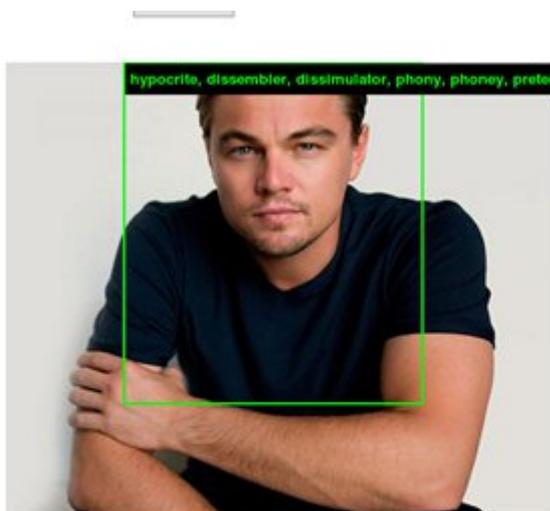theosophist

While I have come across a lot of AI and AI-related applications making a positive impact in the real world, I'm afraid ImageNet Roulette is certainly not one among them. Rather, it exposes the dark part of AI. The model is currently being showcased as part of the Training Humans exhibition by Trevor Paglen and Kate Crawford at the Fondazione Prada Museum in Milan.

"ImageNet Roulette is a provocation designed to help us see into the ways that humans are classified in machine learning systems. It uses a neural network trained on the "Person" categories from the ImageNet dataset which has over 2,500 labels used to classify images of people.", describes the creators.

The system uses an open-source Caffe deep learning framework that is trained on the images and labels that belong to the "person" category. ImageNet Roulette first detects faces in the image inputs. If it finds one, the image will be sent to the Caffe model for classification. If not, the original image gets returned along with a label in the upper left corner.

The categories are drawn from WordNet and they contain a lot of offensive categories that might be racist and even misogynistic at times. *"We want to shed light on what happens when technical systems are trained on problematic training data. AI classifications of people are rarely made visible to the people being classified. ImageNet Roulette provides a glimpse into that process – and to show the ways things can go wrong.",* say the creators.

As you can see below, the neural network describes the sample image of Leonardo DiCaprio as "hypocrite", "dissembler", "dissimulator", and a couple more offensive words. The worst part, however, is the description it provided: *"a person who professes beliefs and opinions that he or she does not hold in order to conceal his or her real feelings or motives"*.

e, dissembler, dissimulator, phony, phoney, pretender: *a person who professes beliefs and opinions that he or she does not hold in order to conceal his or her real feelings or*

rson, individual, someone, somebody, mortal, soul > bad person > wrongdoer, offender > deceiver, cheat, cheater, trickster, beguiler, slicker > hypocrite, dissembler, dissimulator, ony, phoney, pretender

In fact, ImageNet Roulette sets an ideal example of what an AI model should not be and how worse it could get if the wrong dataset is used for training AI-based models. So, what are your thoughts on ImageNet Roulette? Let us know in the comments.

For Original Post Please Go To

[ImageNet Roulette Shows How AI Stereotypes Humans](#)

Loading...