

The Data Behind Deep Fakes

A Deepfake is a photo, video or audio file that has been digitally altered or created to misrepresent reality [1]. In this talk I will describe the consequence of Deepfake technology, new challenges and propose a way to describe training datasets.

Last summer, I joined Reality Defender [2], working closely with the Chief Defender and learned how hard it is to detect Deepfakes even using machine learning. Since the term was coined in 2017, Deepfakes have improved tremendously where we may no longer be able to tell the difference between fakes and reality.

If we train an AI to detect Deepfakes, the detection can only be as good as the suitability of the training data set for the challenge it is facing during inference...

[1] Deepfakes, Explained, Meredith Somers, MIT Sloan:

<https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>

[2] <https://www.realitydefender.com/>

a
of
a

AI



Benjamin Mencer is a part time employee at Reality Defender (www.realitydefender.com) and second-year CS undergraduate at the University California, Santa Cruz. Last year, Benjamin gave talk at the AI seminar at Stanford SLAC.

Before UCSC, Benjamin was inspired to pursue by attending a summer course taught by Ross Alexander at Stanford and lectures by Andrew Ng. Beyond AI, Benjamin is an avid basketballer, having

played in the National League U19 in the United Kingdom.