Assessing Deepfake Video Detection Ability Using Unsupervised Machine Learning

Zachary Tidler
Georgia Institute of Technology, Atlanta, Georgia, United States

Thereza Pereira
Federal University of Paraiba, Rio Tinto, Brazil

Gernel Lumacad
St. Rita's College of Balingasag, Balingasag, Philippines

Richard Catrambone
Georgia Institute of Technology, Atlanta, Georgia, United States

Abstract

As techniques for generating deepfake videos (in which a generative adversarial network is used to alter the human subject of a video) improve, so does the potential to use them for nefarious ends. It is therefore important that we identify those among us who are most susceptible to being fooled. Tidler and Catrambone (2021) presented results which suggest that one's political orientation and affect detection ability are strong predictors of one's ability to detect deepfakes. The current work is a re-analysis of their data, intended to reveal a more highly-resolved profile of a susceptible individual. The data (N = 169) were submitted to a k-means clustering algorithm which revealed a 2-cluster solution.

Younger, more progressive thinking individuals, with agnostic or atheistic religious beliefs, and greater adeptness at spotting deepfakes formed Cluster 1; and individuals with more conservative political leanings, greater religiosity, and less skill at detecting deepfakes formed Cluster 2.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.