SaTC: Core: Small: Decentralized Attribution and Secure Training of Generative Models (# 2101052)

Challenge 1: No *provable* method for attributing an ever-growing number of models w/o a *centralized* classifier.

Challenge 2: No analysis on the trade-off among attribution, generation quality, and the capacity of attributable models.

Challenge 3: Scalable secure training cannot be achieved due to limitations of secure computation on encrypted values and nonlinear functions.

Contribution 1: Sufficient conditions and effective algorithms for certifiable decentralized attribution.

Contribution 2: Private join and compute (PJC) for secure computing over multiple databases, with application to privacy-preserving training of generative models.

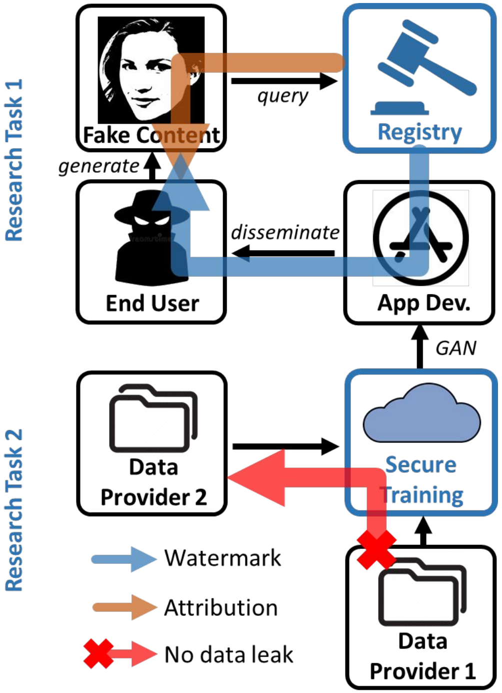
Project number: 2101052
Institution: Arizona State University
Contacts: Yi Ren, Ni Trieu, 'YZ' Yezhou Yang
(yiren, nitrieu, yz.yang@asu.edu)

Damon B. and Marcus G. (2018, Feb 2). A guide to "DeepFakes", the internet's latest moral crisis

James V. (2020, Jun 23). **What a machine learning tool that turns Obama white can (and can't) tell us about AI bias**

Objective 1: Attribute generated contents to their source models correctly.

Objective 2: Prevent data leak in collaborative training among data providers.



Scientific Impact 1: Connecting model attribution with the open challenge of optimal packing of non-convex objects in a high-dim space.

Scientific Impact 2: New functionality – Private Join and Compute. New cryptographic primitive – private information retrieval (PIR) with default.

Broader impact 1: Social - Address threats from malicious personation (generative DeepFake) and biased data/model applications.

Broader impact 2: National security - Secure training on private data (e.g., for collaboration among manufacturers).

Broader impact 3: Education and outreach - Cross-disciplinary course materials on cyber-security, machine learning, and optimization theory.