WashU

Office of Technology Management

ANTIFAKE: COMPUTER METHOD DESIGNED TO PREVENT UNAUTHORIZED SPEECH SYNTHESIS

Yu, Zhiyuan, Zhang, Ning

Maland, Brett

T-020985

Published Date: 2/14/2025

Value Proposition: *Speech synthesizer that uses adversarial audio to prevent the creation of deepfakes.*

Technology Description

Researchers at Washington University in St. Louis have developed a computer implemented method for generating audio content that introduces perturbations into regular speech so that AI models are not able to use the content to generate synthesized speech. Current technologies that uses deep neural networks and generative AI has catalyzed growth in realistic speech synthesis, leading to the emergence of "DeepFake" where synthesized speech can be misused to deceive humans and machines for nefarious purpose.

This method, labeled AntiFake, utilizes a defense mechanism that relies on adversarial examples to prevent unauthorized speech synthesis, ensuring the transferability to attackers' unknown synthesis models by using an ensemble learning approach to improve the generalizability of the optimization process.

Stage of Research

Tested against market leading synthesizer and five other black box synthesizers for 24 participants

Publications

Yu, Zhiyuan, Zhai, Shixuan, & Zhang, Ning. AntiFake: Using Adversarial Audio to Prevent Unauthorized Speech Synthesis. Proceedings of the ACM Conference on Computer and Communications Security, (). https://doi.org/10.1145/3576915.3623209

Applications

- Speech Synthesizer

Key Advantages

- Can be used on multiple synthesis models
- Segment based optimization strategy for perturbations that confuses the speech synthesizer
- Uses frequency penalties based on inverse sound pressure levels

- Prevents the creation of deepfakes

Patents

Patent application filed

Related Web Links – Ning Zhang Profile; Zhang Lab