* Phishing: “practice of sending emails appearing to be from reputable sources with the goal of influencing or gaining personal information.” [n. 网络仿冒，网络钓鱼]
* Vising: [n. 网络电话诈骗]

How to defend Deepfake Audio

Problem

* Imitate person’s speech, generate believable voice clips using only second long samples[1]
* Harder to detect with ears[1]
* Attackers infiltrate the organization/company/account by having the user click on malicious link (email, social media direct message, text message)[2]
* Fake news, empowering criminals, super-change political mayhem, spam calls and whitle-collar crime[4]

Tech-Detection

* AI-based detection methods : identify fakes: as detection gets better, so will the fakes[1]
* Run content through a deepfake detector[3]
* Train computer for inaudible hints that the voice couldn’t have come from an actual person[4]
* **Google recently made** a [vast dataset of its own synthetic speech](https://www.blog.google/outreach-initiatives/google-news-initiative/advancing-research-fake-audio-detection/) available to researchers who are working on deepfake detection. This trove of training data can help AI systems find and recognize the hallmarks of fake voices.[4]

Pindrop

* Pindrop: an Atlanta company that sells voice authentication to big banks and insurance companies, is also developing defenses, worried that the next wave of attacks on its clients will involve deepfake audio.[4]
* One key to detecting fakes, according to the company: sounds that seem normal, but that people aren't physically capable of making.[4]
* An example from Pindrop CEO Vijay Balasubramaniyan: If you say "Hello, Paul," your mouth can only shift from the "o" to "Paul" at a certain speed. Spoken too fast, "the only way to say this is with a 7-foot-tall neck," Balasubramaniyan says.[4]
* Pindrop, the audio biometrics company, is developing synthetic voices in order to [train its own defenses to detect them.](https://www.axios.com/deepfake-audio-ai-impersonators-f736a8fc-162e-47f0-a582-e5eb8b8262ff.html)[5]

Labelling

* The videos/voice should be labeled is something is detected as being manipulated[3]
* hold those platforms who host and make deepfakes available to the public accountable and responsible for them[6]
* If a post has not had any type of trusted source or context provided, then correct labeling of the content should be clear to the viewer that the content source has been verified, is still being analyzed, or that the content has been significantly modified.[6]

Raise user awareness (Company)

* Enforce strict verification procedure[1]
* Practice verification when recognize someone’s voice or face[1]
* Educate employees for the verification procedure[1]
* Train employees to properly answer, identify and react to suspicious calls[2]
* Engineers can develop and implement security training program[2,6]
* make sure your security training includes identifying modern tactics like deepfakes and mobile phishing – especially while people work remotely. Since we can’t walk down the hall to validate communication from a co-worker, encourage your employees to reach out over different channels.[6]
* sending a message through a collaboration system to verify that an unusual phone call was legitimate.[6]

Secure digital profile (Organization/Individual)

* Keep software up to date (security updates)[2]
* Routinely check privacy settings[2]
* Strong password[2]
* Account security: Multi-factor authentication on every account[1]
* Never navigate to a website from phone call or email link (check validity)[2]
* Check validity of callers (Solution: request the caller’s email address to send a email to check their identity, it can be similar to what websites has done, after entering correct password, they would send a security code though text, phone call or email)[2]
* Look up caller though employee system[2]

Individual awareness in data protection

* Limiting public presence on social media[1]
* Enabling privacy restrictions[1]
* Prevent scammers form easily stealing your voice[1]

Future Problem:

As detection gets better, so will the fakes[1]

Reach to a point where it is not going to be possible to detect AI fakes at all[3]

* The videos/voice should be labeled is something is detected as being manipulated[3]
* Frictionless sharing and monetarization of attention[3]

How to evaluate the effectiveness of such defense

* Enforce verification procedure and educate employees: [1]

Test employees by having them receive live calls from trained professionals who can emulate the tactics of real attackers[1]

What does it mean to audio/voice anonymization

* Relation between voice anonymization and deepfake audio defense
* Are they the same/

Slides

Reference

[1] Seorg, “Deepfakes: How to defend yourself from attack,” Security Boulevard, 20-May-2020. [Online]. Available: https://securityboulevard.com/2020/05/deepfakes-how-to-defend-yourself-from-attack/.

[2] Social-Engineer, “Secure it – keep your digital profile safe from Vishers and phishers,” Social, 20-Nov-2020. [Online]. Available: https://www.social-engineer.com/secure-it-keep-your-digital-profile-safe-from-vishers-and-phishers/.

[3] J. Vincent, “Deepfake detection algorithms will never be enough,” The Verge, 27-Jun-2019. [Online]. Available: https://www.theverge.com/2019/6/27/18715235/deepfake-detection-ai-algorithms-accuracy-will-they-ever-work.

[4] K. Waddell, “Researchers are figuring out how to detect audio deepfakes before it's too late,” Axios, 03-Apr-2019. [Online]. Available: https://www.axios.com/2019/04/03/deepfake-audio-ai-impersonators.

[5] K. Waddell, “This ai-generated Ellen DeGeneres Voice is the future of deepfakes,” Axios, 03-Apr-2019. [Online]. Available: https://www.axios.com/2019/04/03/deepfake-audio-ai-ellen-degeneres.

[6] C. Franklin, “Defending against deepfakes: From tells to crypto,” Dark Reading, 17-Sep-2020. [Online]. Available: https://www.darkreading.com/edge/defending-against-deepfakes-from-tells-to-crypto.