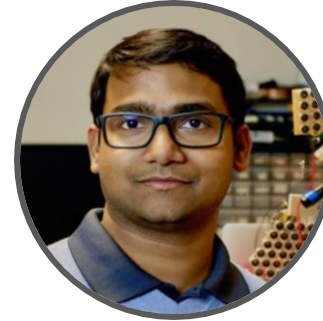


For Human Ears Only: Preventing Automated Monitoring on Voice Data



Irtaza Shahid
irtaza@umd.edu



Nirupam Roy
niruroy@umd.edu

Audio-based Communication is everywhere



Speech contain rich information ...



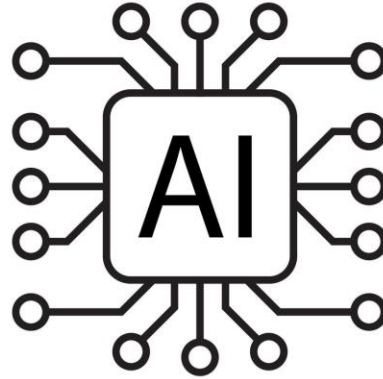
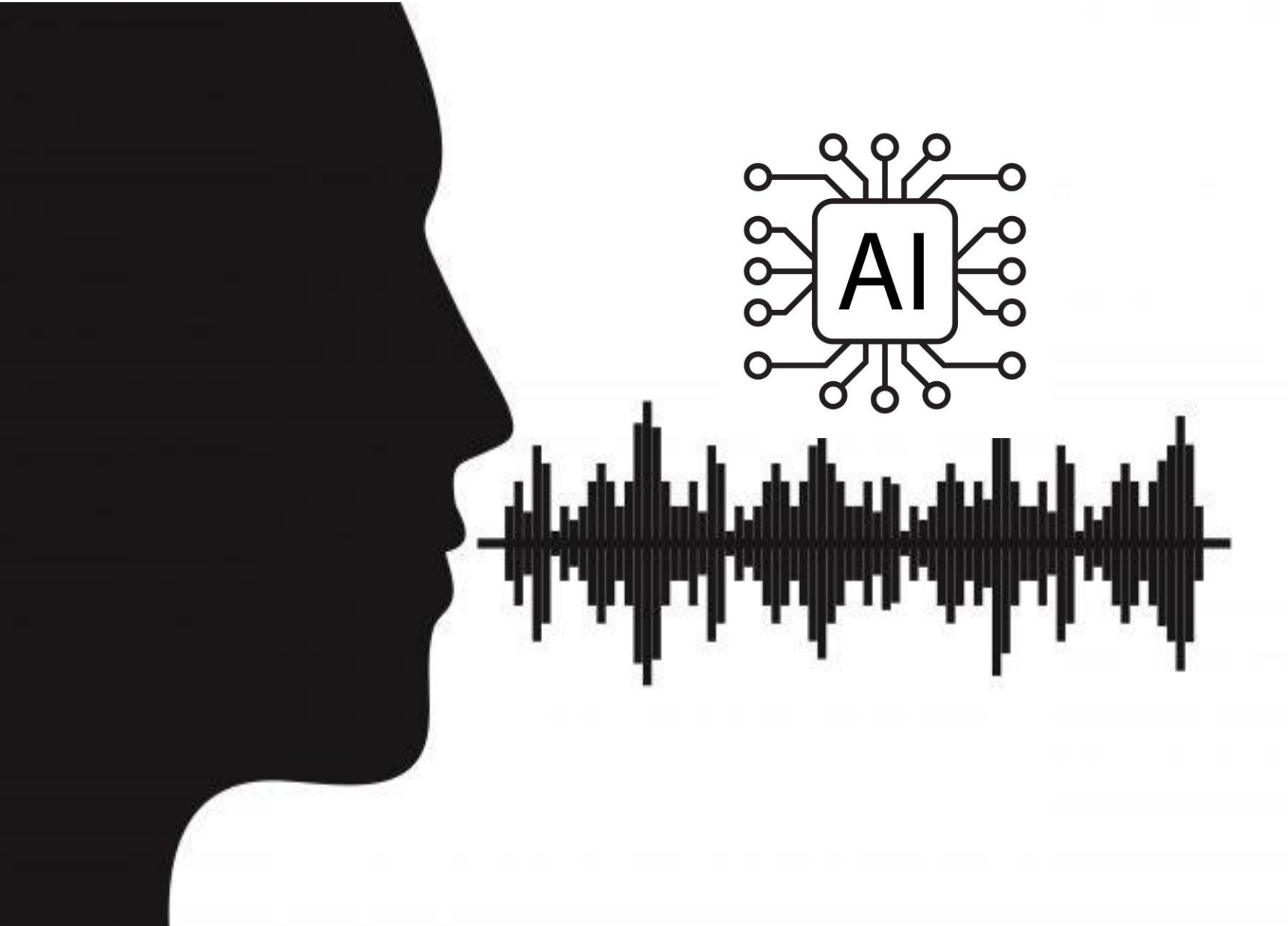
- Speech Content

Speech contain rich information ...



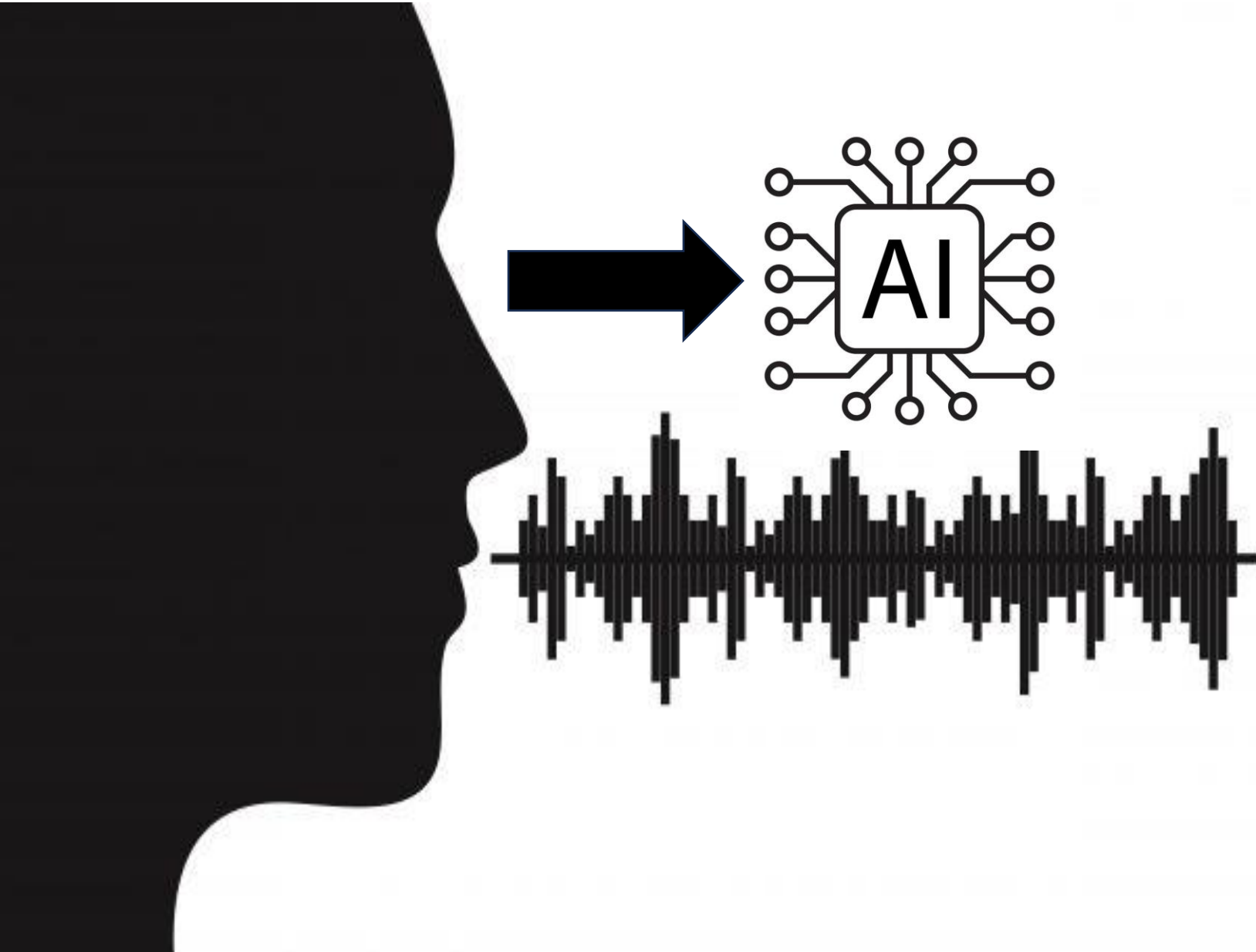
- Speech Content
- Speaker Identity
- Gender
- Geography
- Age
- Emotional cues
- Health signatures

Easier to extract information



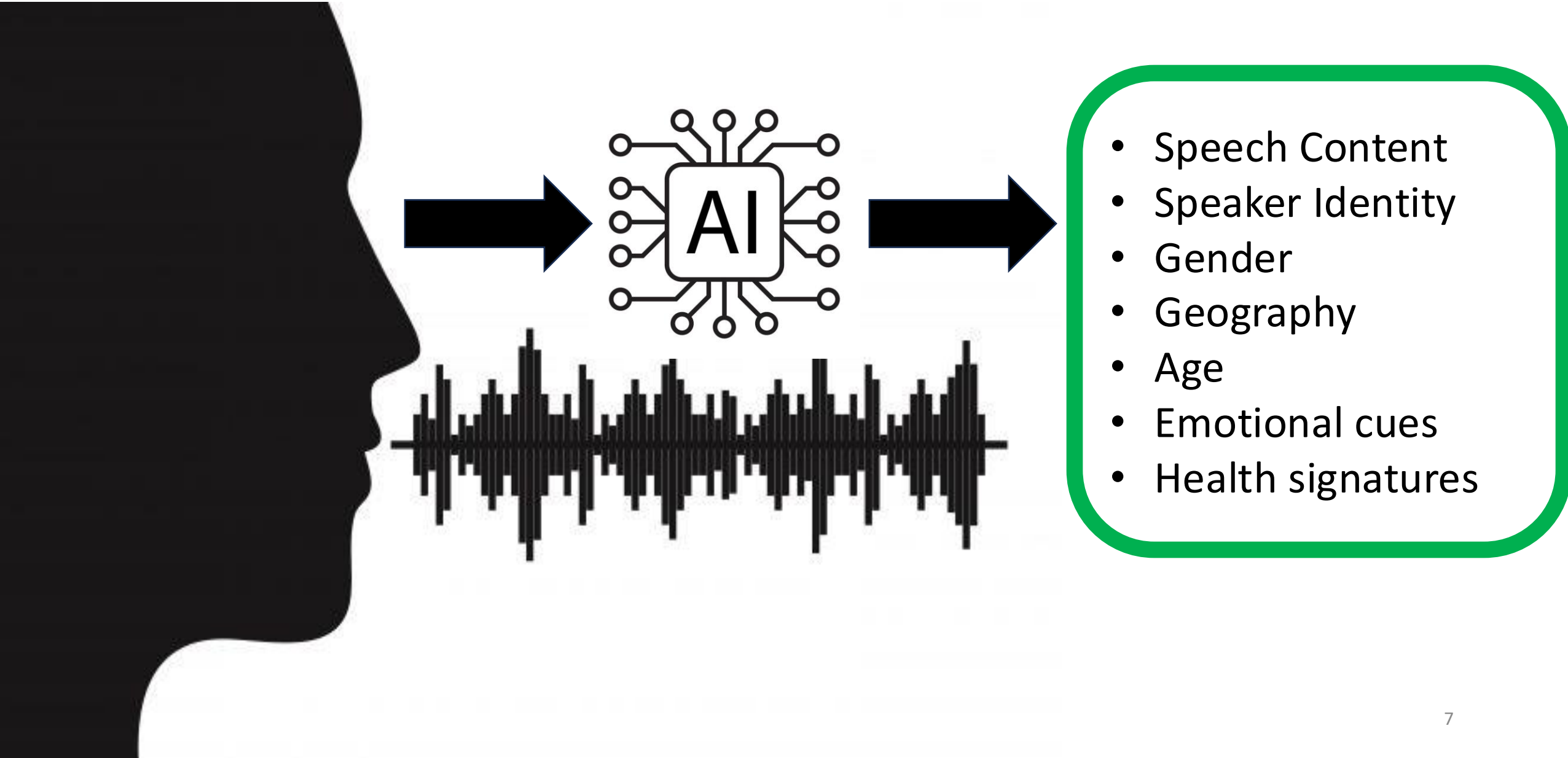
- Speech Content
- Speaker Identity
- Gender
- Geography
- Age
- Emotional cues
- Health signatures

Easier to extract information

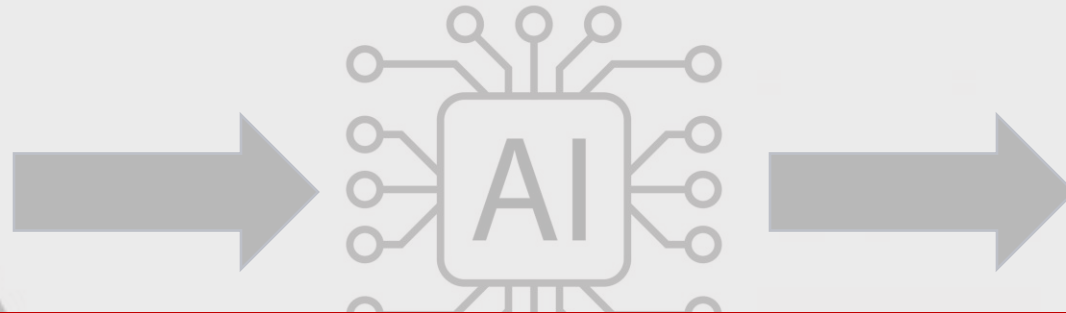


- Speech Content
- Speaker Identity
- Gender
- Geography
- Age
- Emotional cues
- Health signatures

Easier to extract information



Easier to extract information



- Speech Content
- Speaker Identity
- Gender

Today over 95% data is transferred digitally

- Emotional cues
- Health signatures

If get accessed ...



Mass Surveillance

If get accessed ...

Mass Surveillance

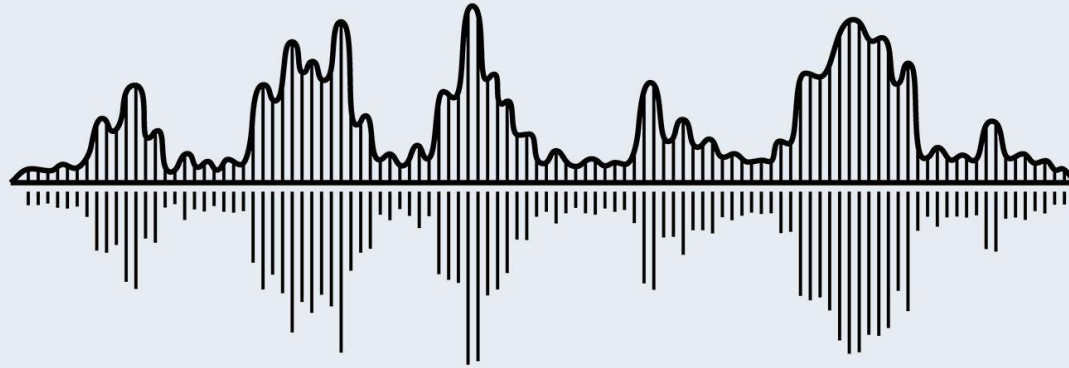


Stealing personal information

If get accessed ...



Mass Surveillance

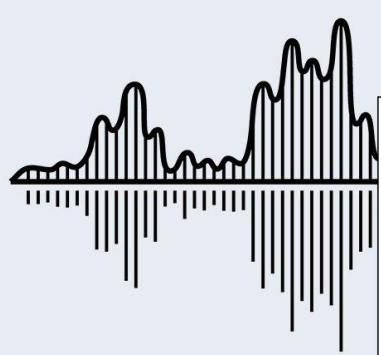


Ai Voice Cloning

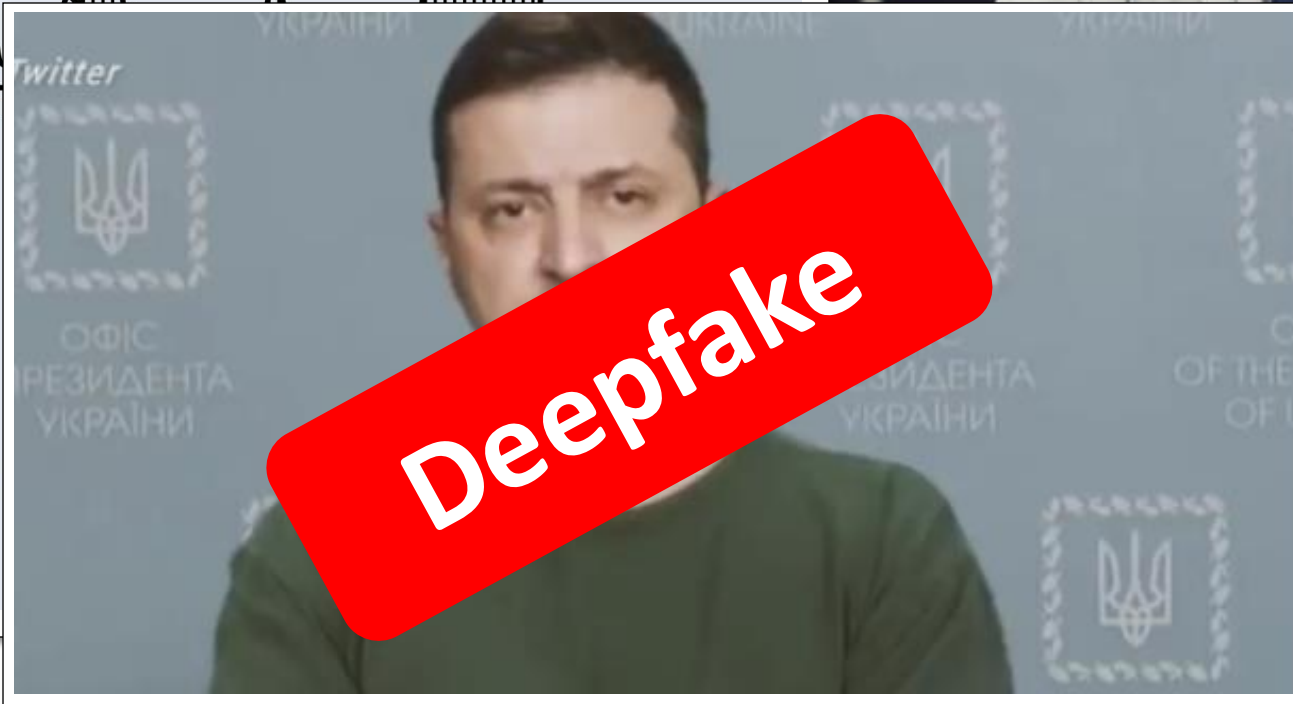


Stealing personal information

If get accessed ...

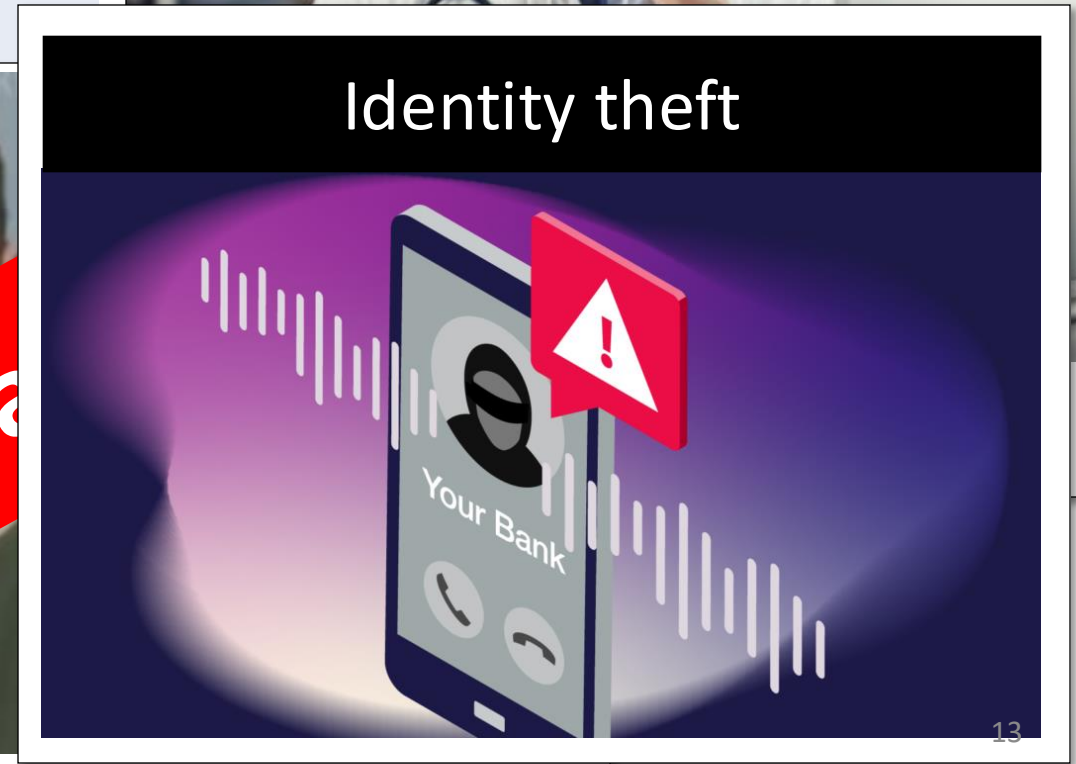
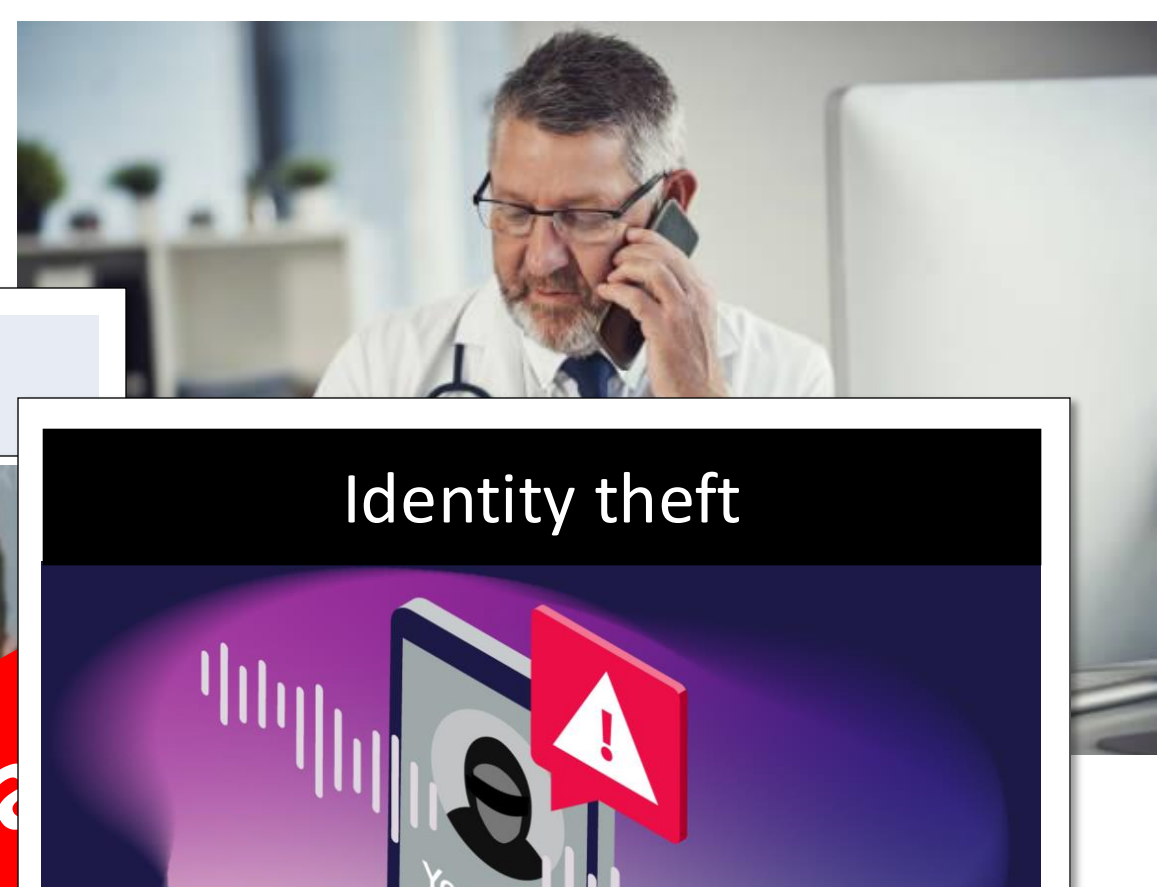


Ai Vo



al information

If get accessed ...



Threat is real ...



Nearly all AT&T cell customers' call and text records exposed in a massive breach

Threat is real ...



Nearly all AT&T cell customers' call and text records exposed in a massive breach



Chinese hackers said to have collected audio of American calls

Threat is real ...

Forbes

Fraudsters Cloned Company Director's Voice In \$35 Million Heist, Police Find



CN
Nearly record

Cybercriminals cloned the voice of a company director in the U.A.E. to steal as much as \$35 million in a huge and complex heist. GETTY

The Washington Post
Democracy Dies in Darkness



Chinese hackers said to have collected audio of American calls

Threat is real ...



Forbes
Fraudsters Cloned Company
Director's Voice In \$35 Million
Heist, Police Find

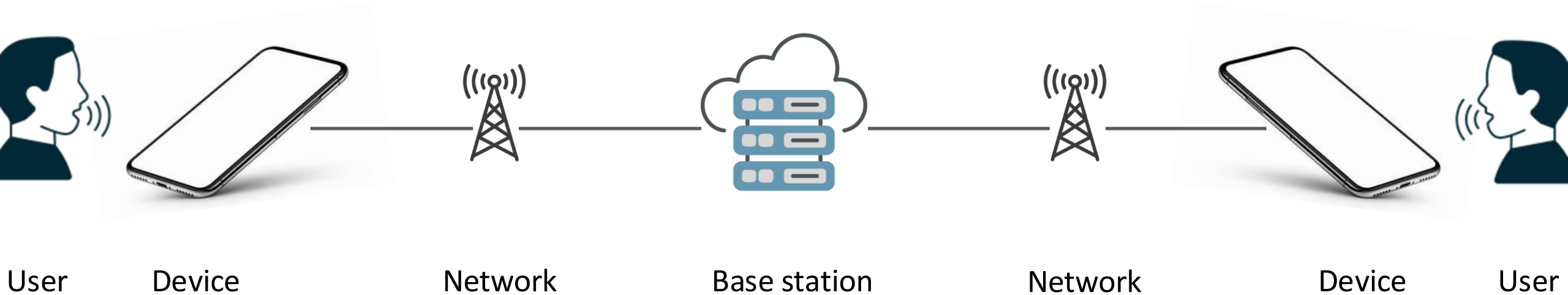


Cybercriminals cloned the voice of a company director in the U.A.E. to steal as much as \$35 million in a complex heist. GETTY

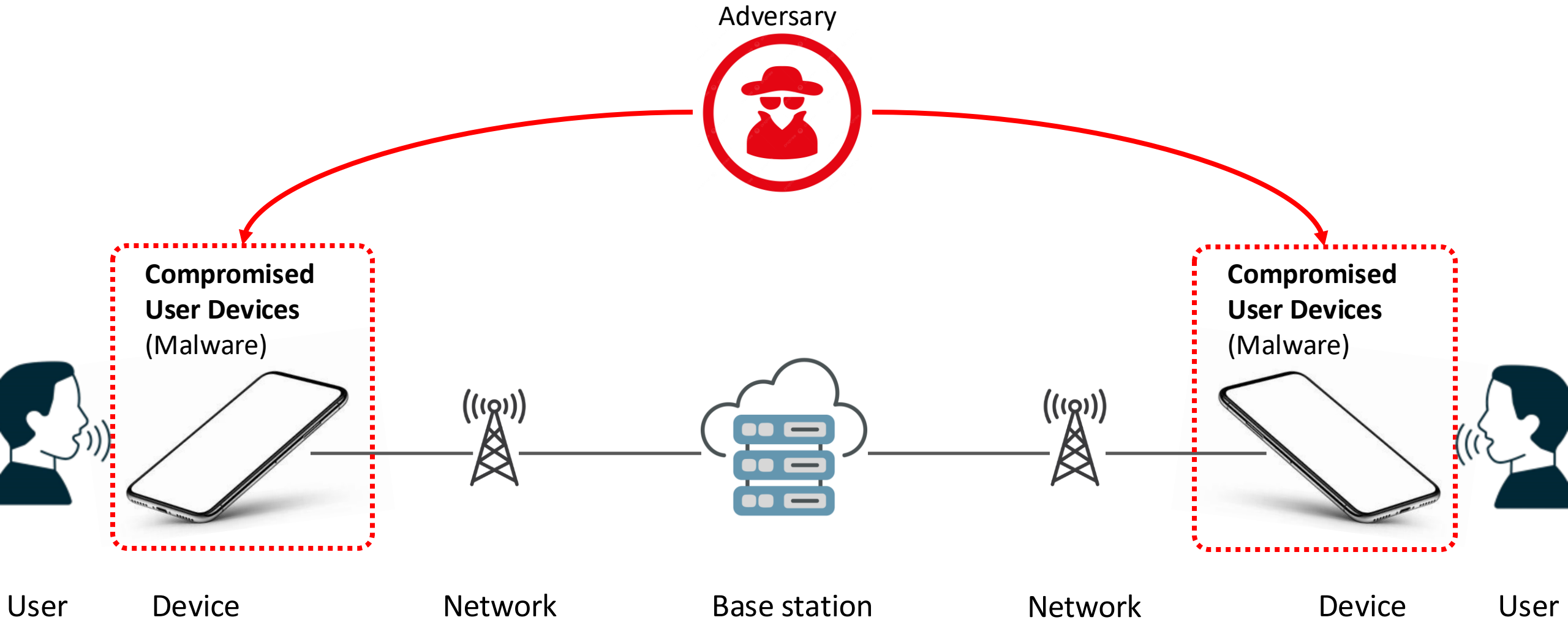


WIRED
Secret Phone Surveillance Tech Was Likely Deployed
at 2024 DNC

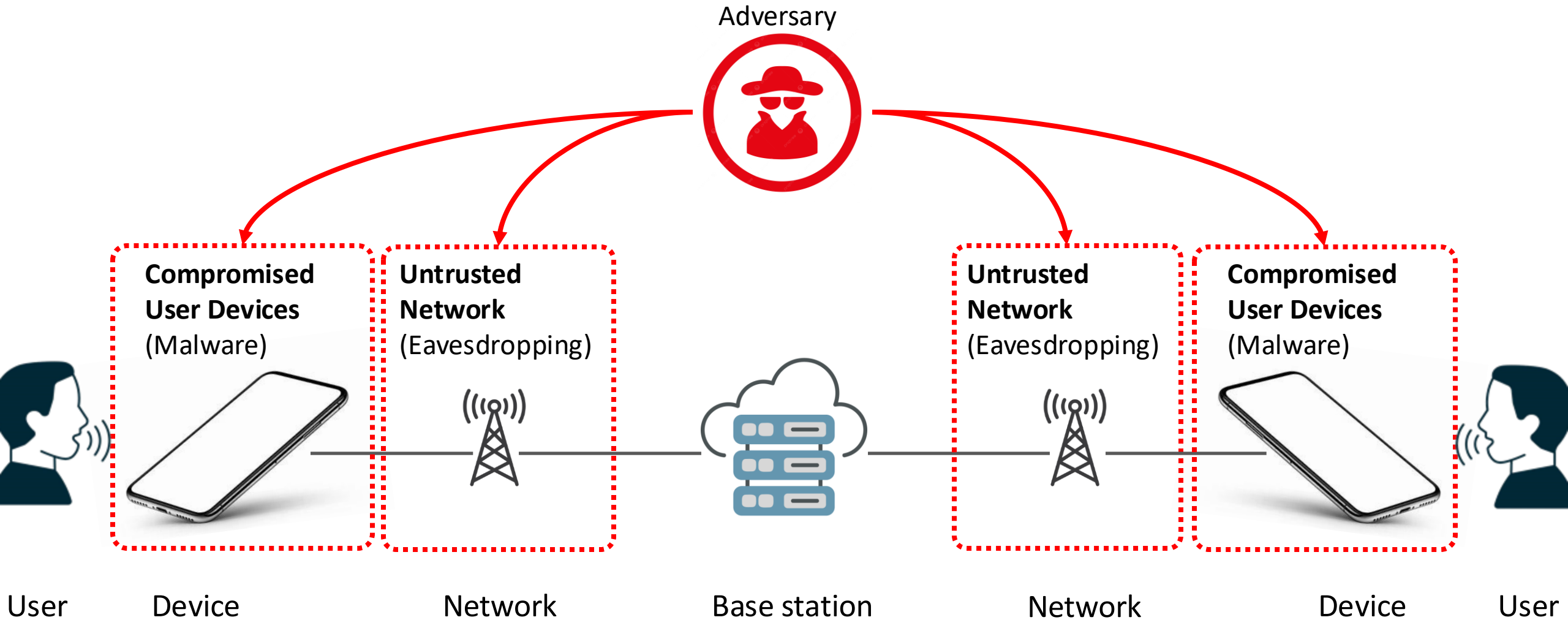
Voice Communication Channel



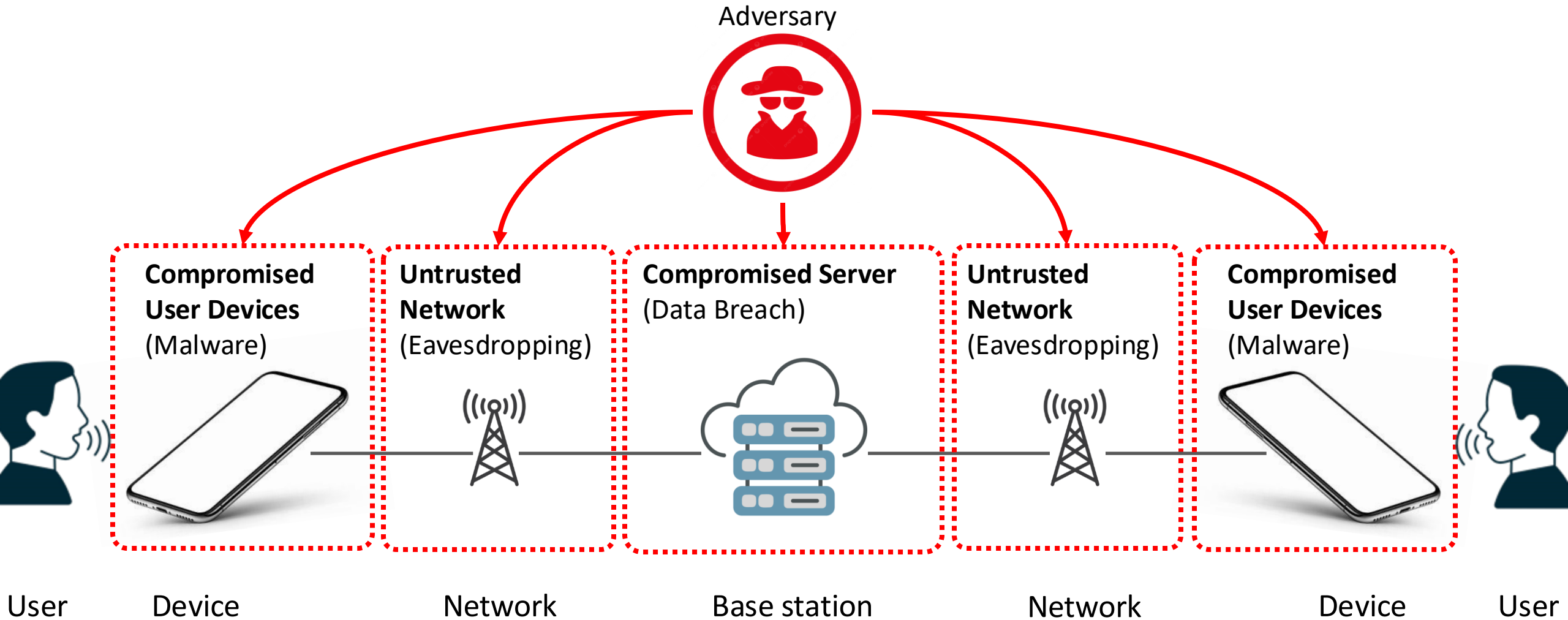
Voice Communication Channel



Voice Communication Channel

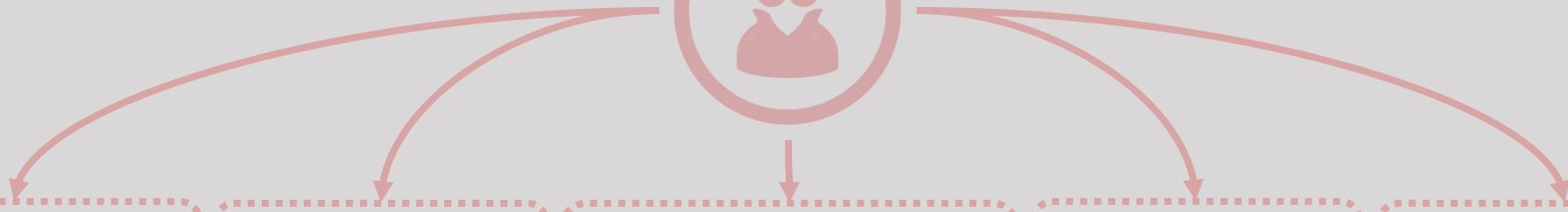


Voice Communication Channel



Voice Communication Channel

Adversary



Is it possible to prevent automated monitoring?



Solution should ...

Prevent automated monitoring

Do not disrupt natural communication

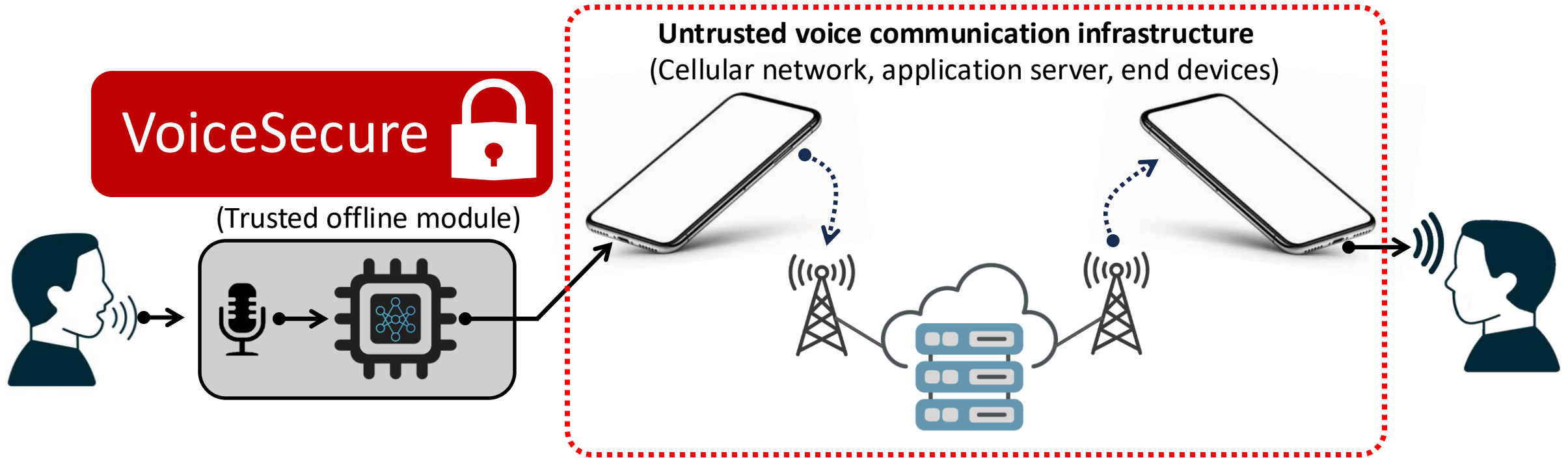
Protect speech before devices

Adversary

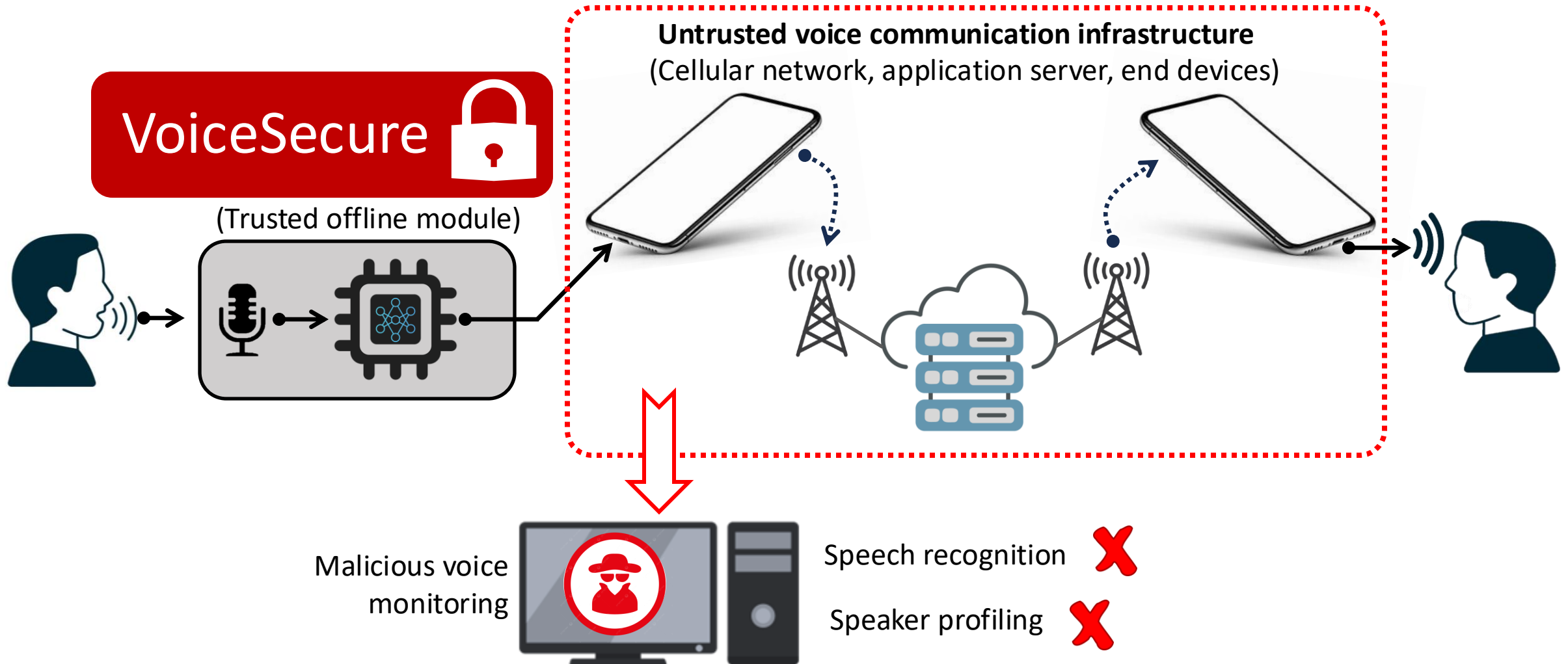
Untrusted
Network
(Eavesdropping)

Compromised
User Devices
(Malware)

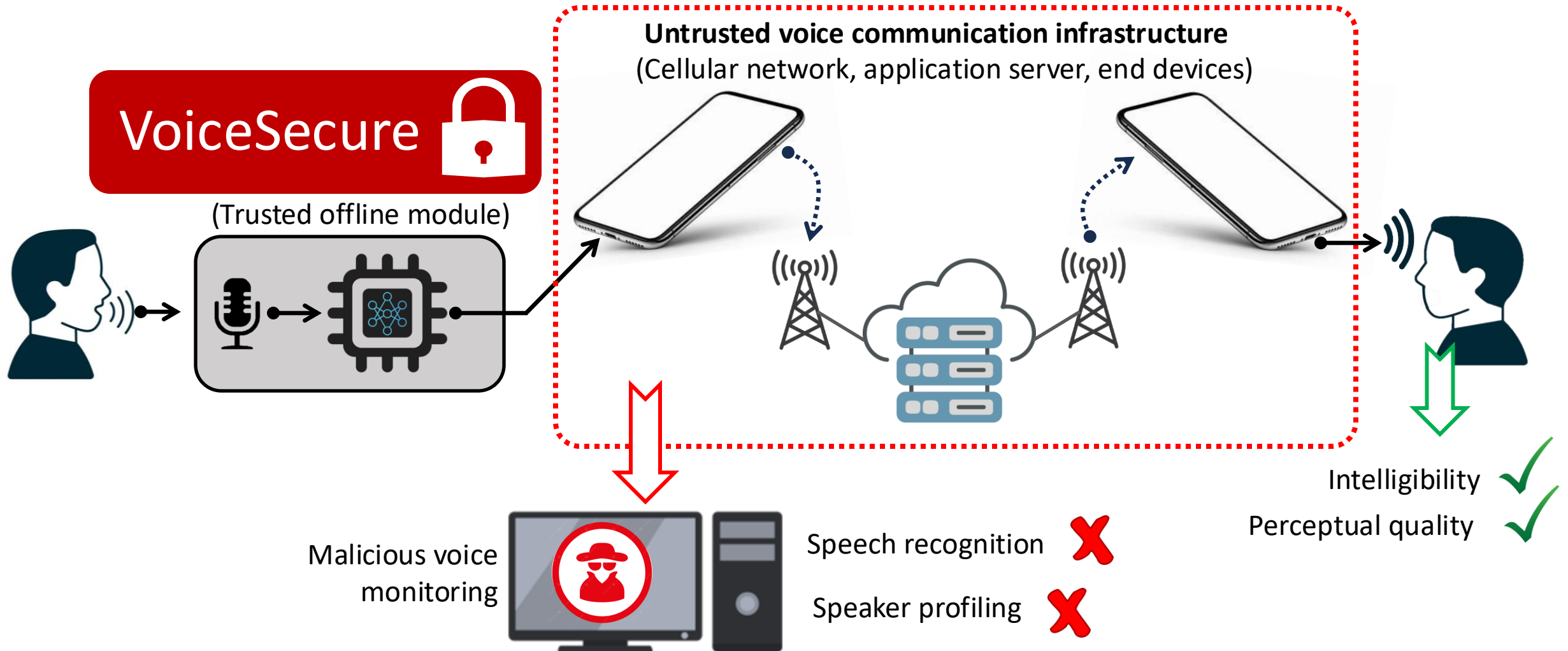
VoiceSecure



VoiceSecure



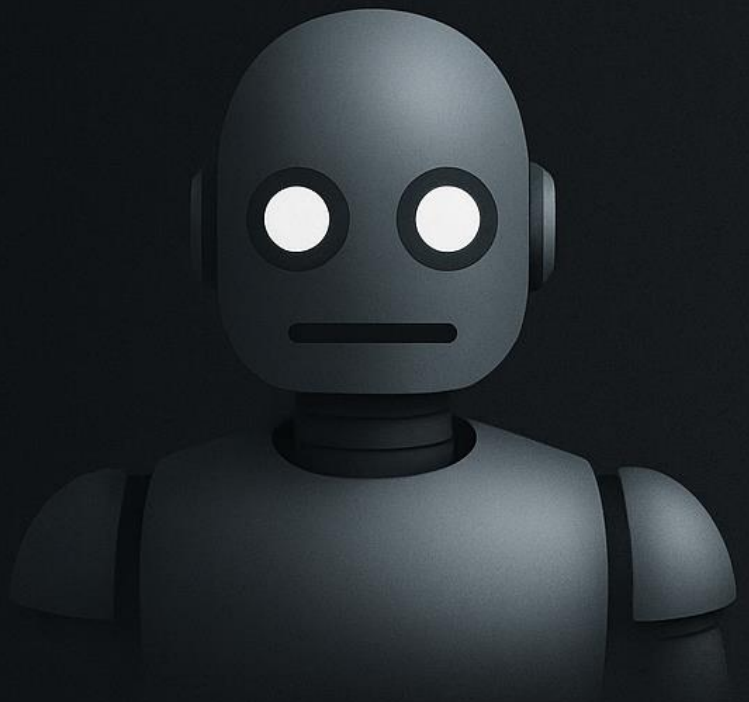
VoiceSecure



Existing Solutions

Method	Speaker Identification	Speech Recognition	Naturalness	Real-time	Hardware
McAdams [1]	✓	✗	✓	✓	✓
VoiceMask [2]	✓	✓	✗	✗	✗
VCloak [3]	✓	✗	✓	✗	✓
SMACK [4]	✓	✓	✓	✗	✗
Stop Bugging Me [5]	✓	✗	✗	✓	✓
MicPro [6]	✓	✗	✓	✓	✓
<i>VoiceSecure (Our Solution)</i>	✓	✓	✓	✓	✓

Machines



0.63 0.15 0.91 0.44



Humans



Machines

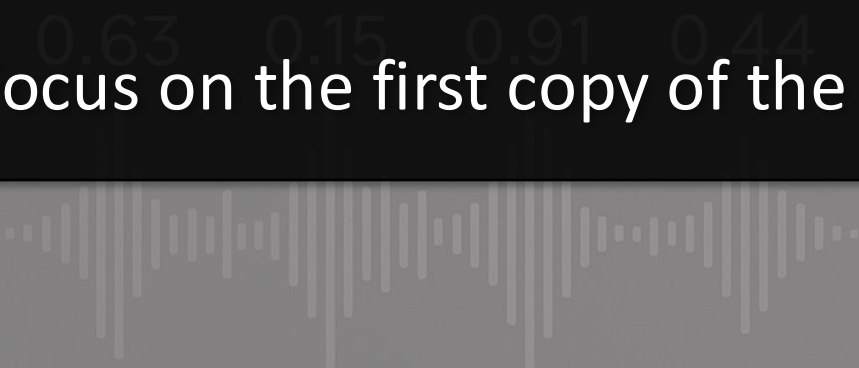
Humans

Humans can tolerate slight variations in fundamental frequencies

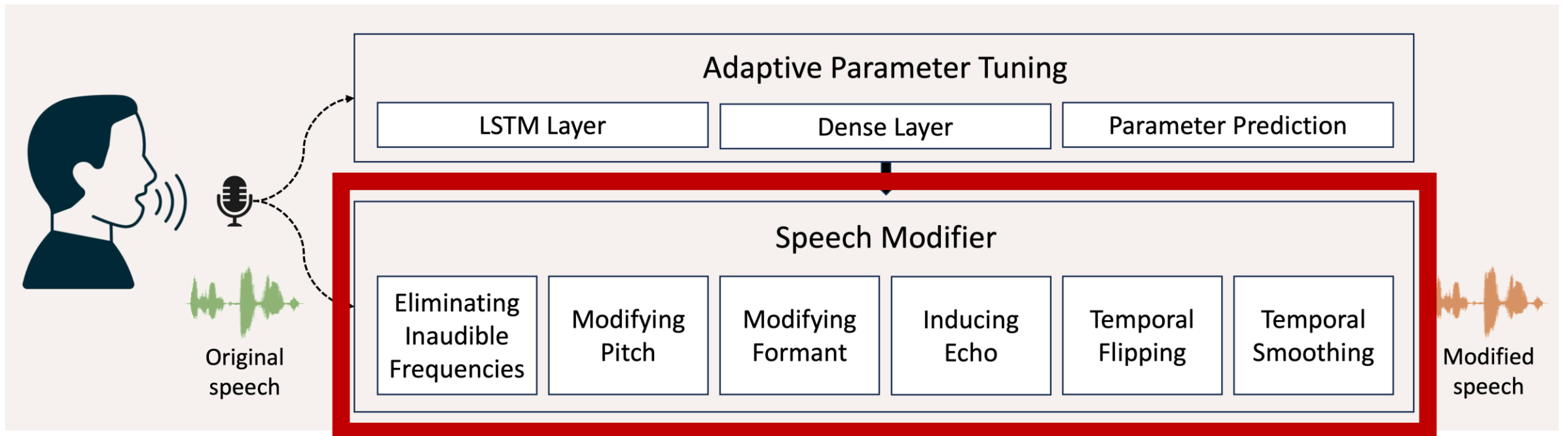
Human brain automatically fill out missing auditory information

Phoneme-sized temporal flipping are imperceptible

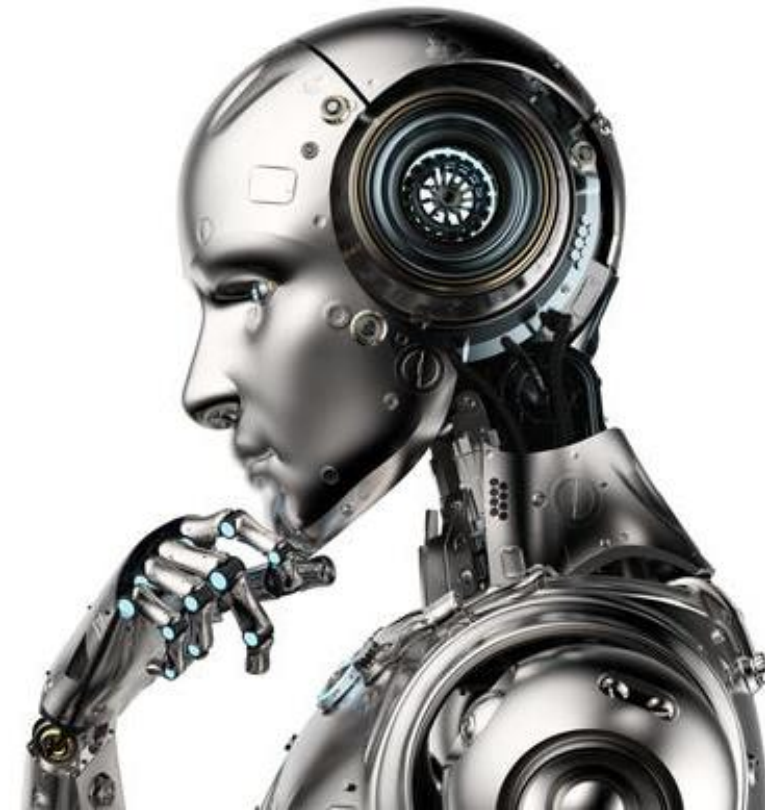
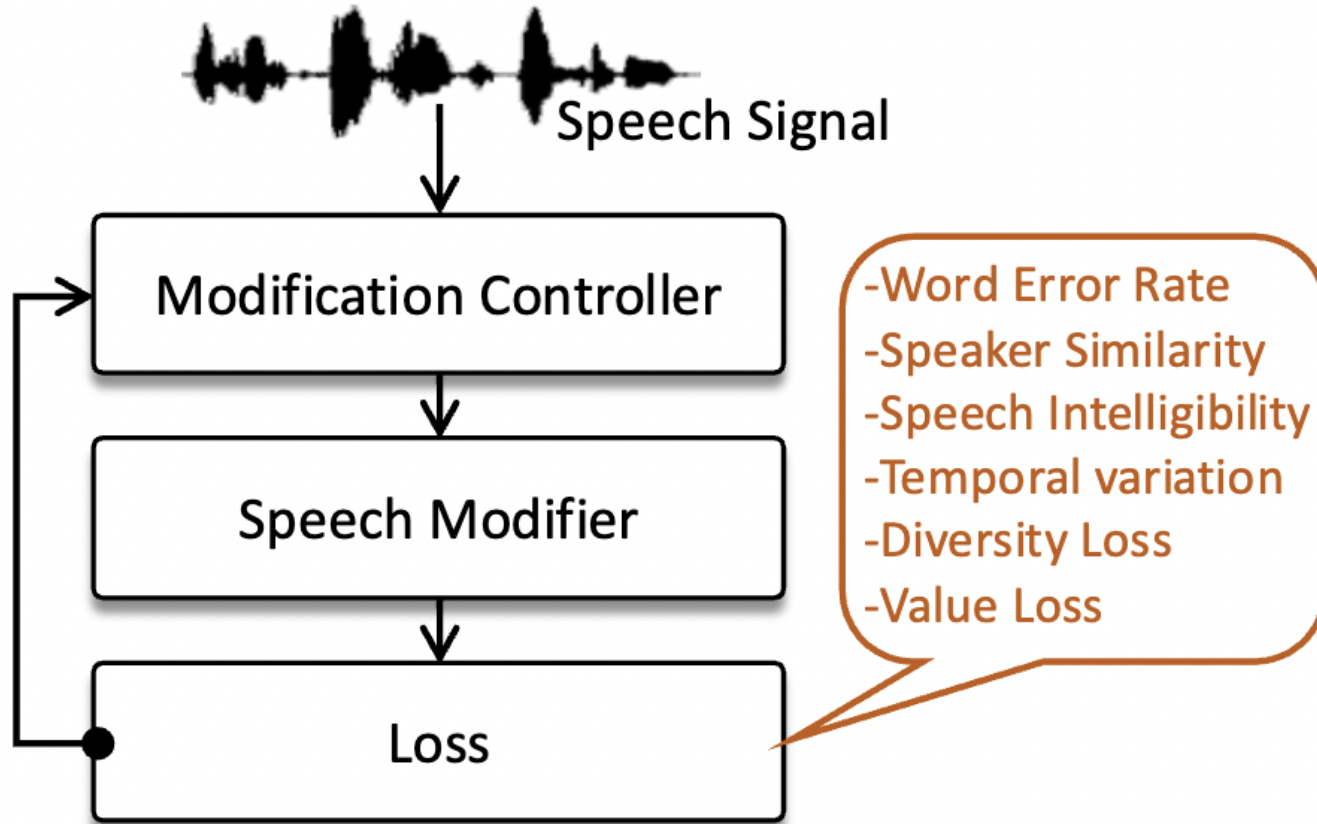
Focus on the first copy of the audio during echo



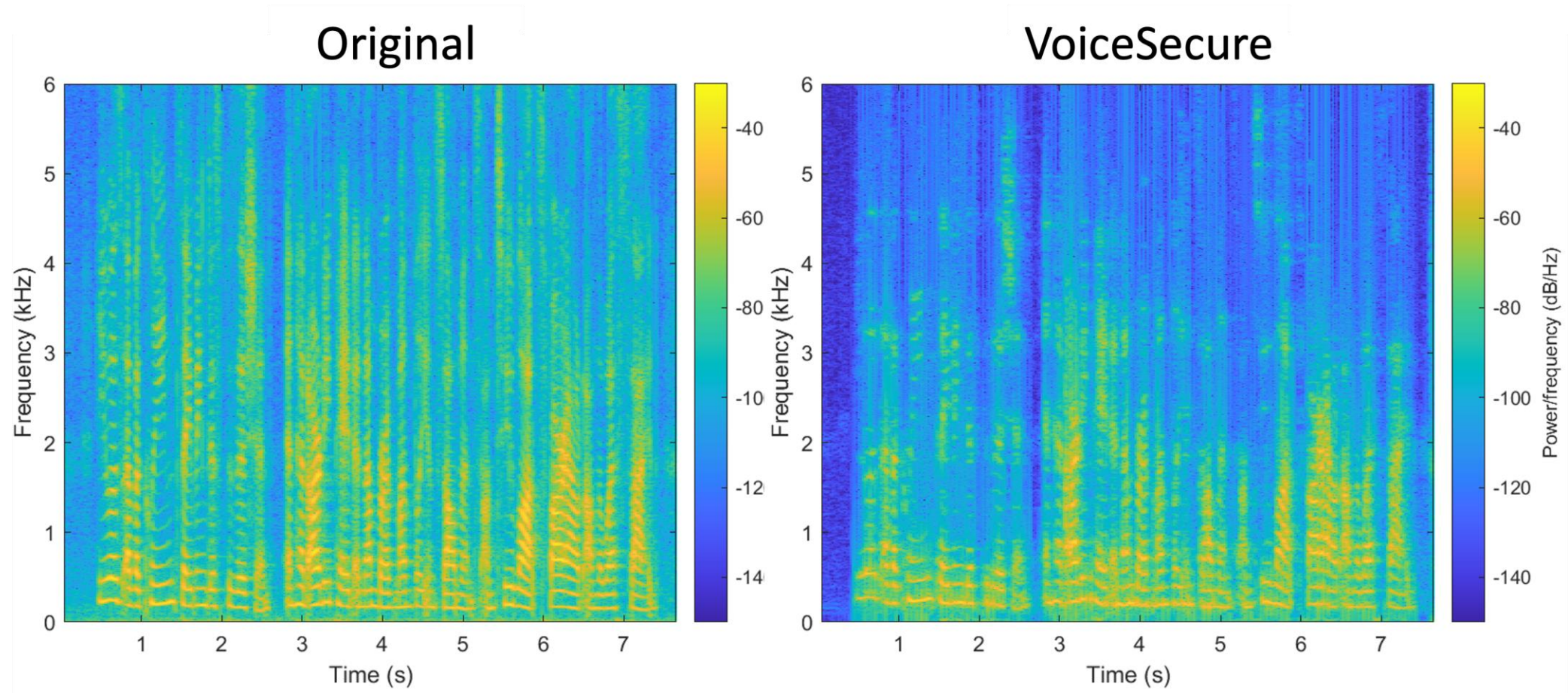
VoiceSecure



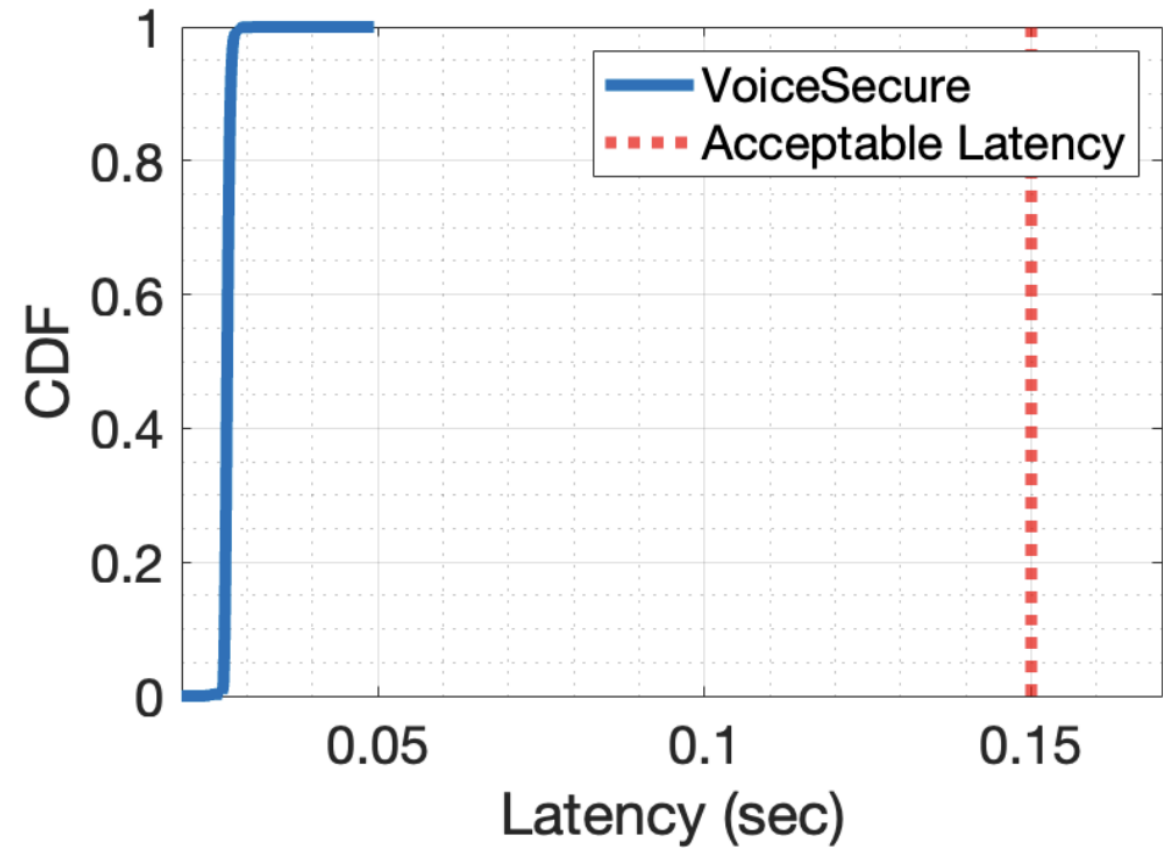
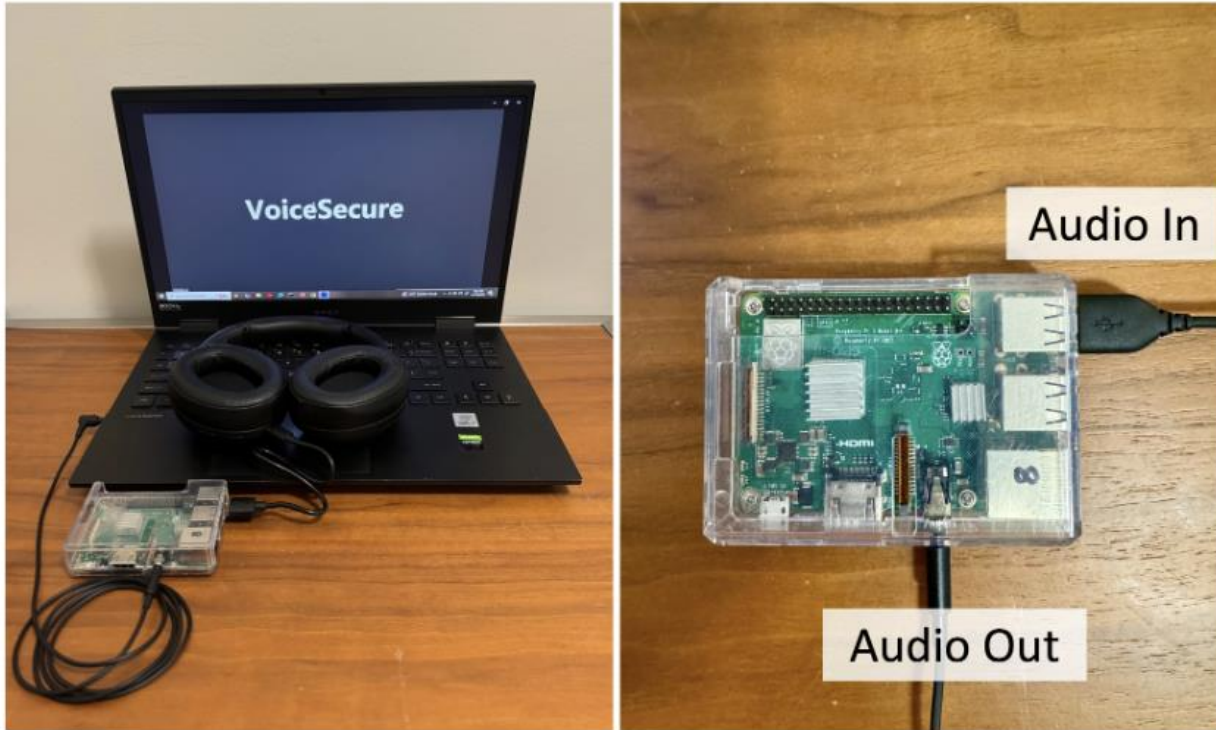
Adaptive Parameter Tuning



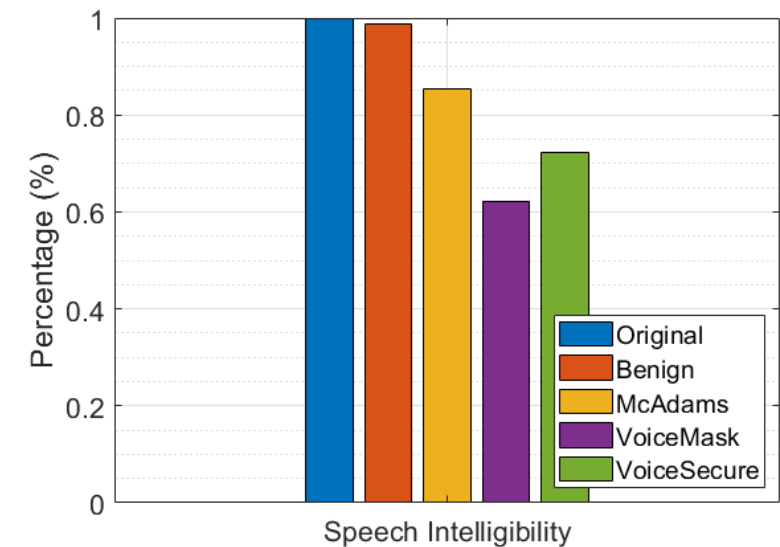
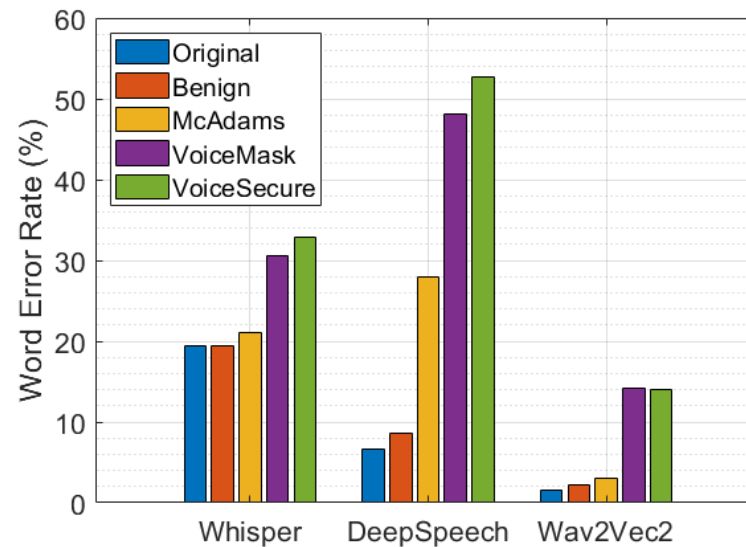
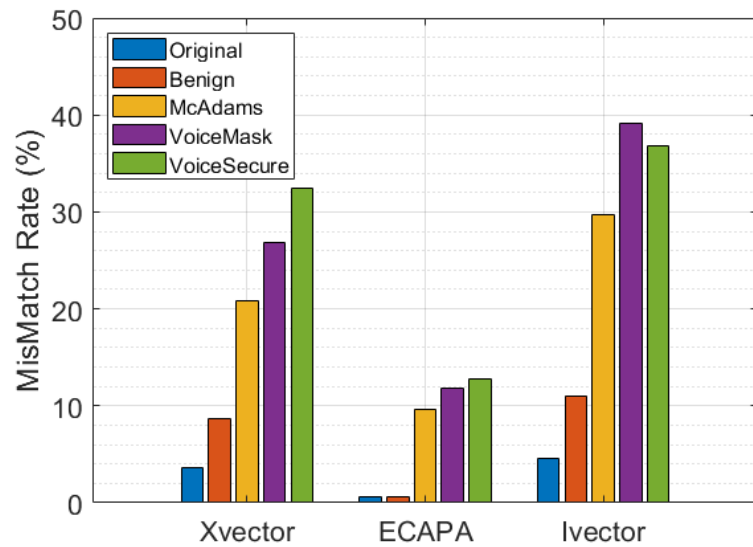
Perceptual Result



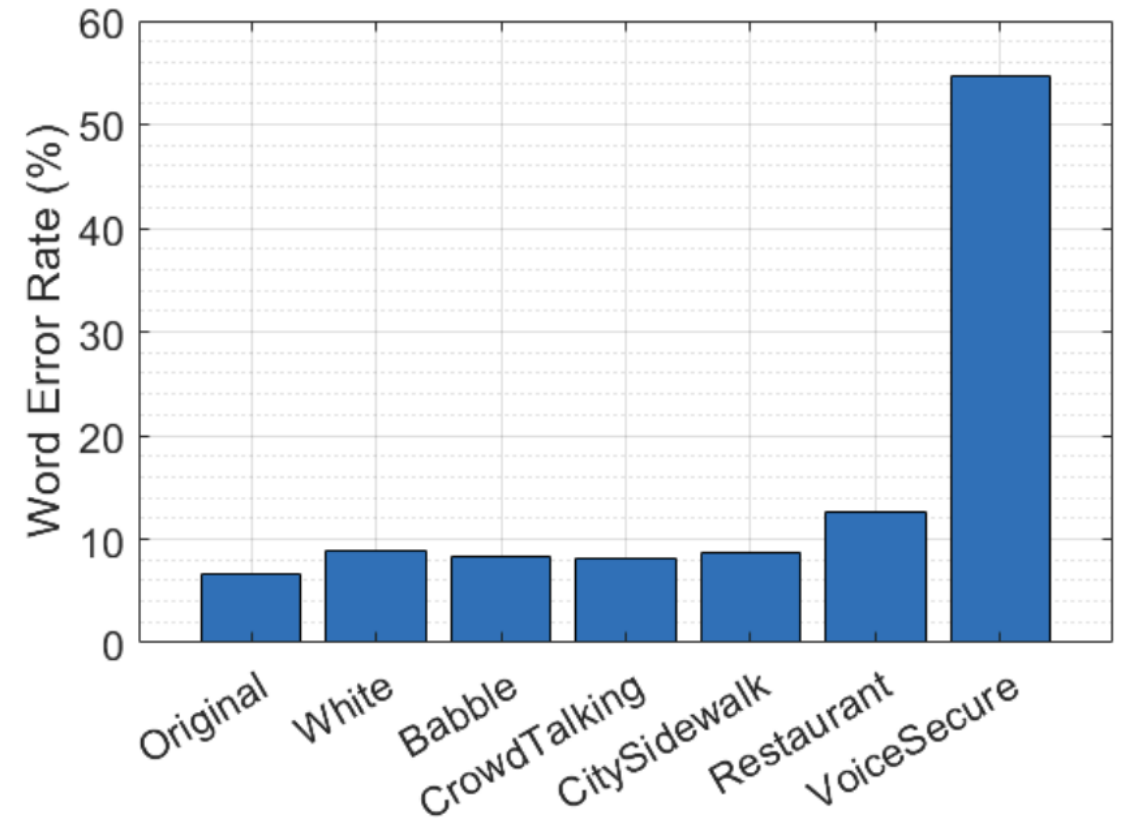
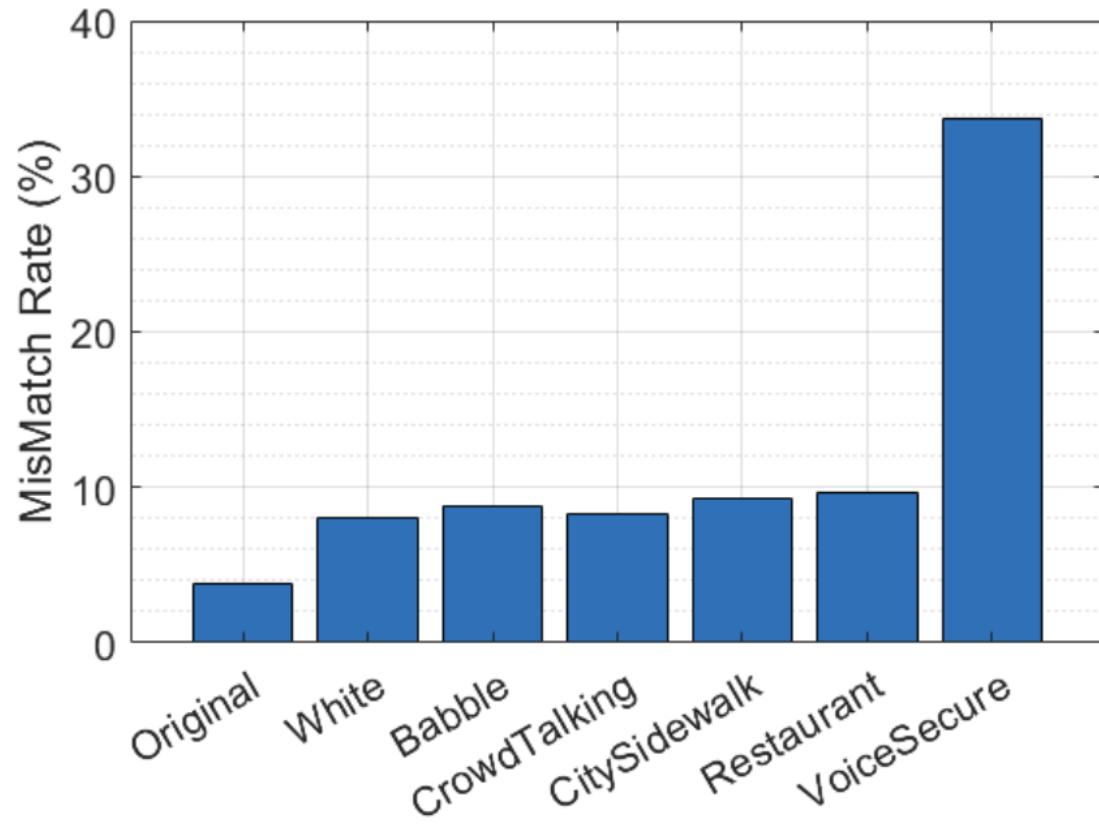
Implementation



Experimentation Results



Experimentation Results



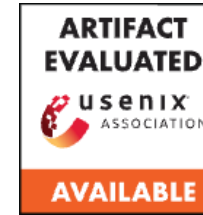


- A microphone module that protect user privacy
- While keeping speech natural
- Real-time processing



UNIVERSITY OF
MARYLAND

34TH USENIX
SECURITY SYMPOSIUM



For Human Ears Only: Preventing Automated Monitoring on Voice Data

Irtaza Shahid, Nirupam Roy

*University of Maryland, College Park
{irtaza, niruroy}@umd.edu*

Abstract

As voice communication becomes an essential part of modern life, the exposure of sensitive information through audio calls presents significant privacy risks. Malicious actors can gain access to this data by compromising user devices, exploiting communication channels, or attacking data servers, making it vulnerable to automated monitoring systems that can identify speakers and extract speech content. To address these privacy concerns, we introduce *VoiceSecure*, the first microphone module designed to prevent automated monitoring of speech while preserving its natural sound for humans. By leveraging the principles of human auditory perception, *VoiceSecure* employs a set of speech modifications that are adaptively tuned in real-time to obscure speaker identity and speech content, without compromising the quality of the audio for human listeners. This hardware-based solution mitigates the risk of software-based attacks, integrating seamlessly with commercial devices via audio jack or Bluetooth. Comprehensive evaluation across state-of-the-art speaker verification and speech recognition models, and a variety of speech datasets, demonstrates that *VoiceSecure* outperforms traditional methods of protecting speech from automated monitoring while keeping it intelligible for humans.

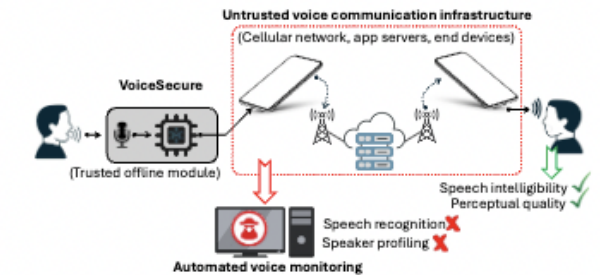


Figure 1: *VoiceSecure*, an offline hardware-software solution to protect speech privacy in real-time voice communication.

Internet Protocol) [4], almost all phone calls, even traditional landline calls, are converted to digital signals for transmission over the internet. This makes the data readily available to computers at various stages of the transmission, starting from the user devices to internet middleboxes, application servers, and even eavesdropping on transmission channels. It has raised concerns about information security. Advancements in voice-to-text technologies and natural language processing algorithms make it possible to launch online monitoring and mass surveillance of digital voice data. Voice is not only a medium for information it carries, it also reveals

Thank you!



Irtaza Shahid

irtaza@umd.edu

<https://www.cs.umd.edu/~irtaza/>



Prof. Nirupam Roy

niruroy@umd.edu

<https://www.cs.umd.edu/~nirupam/>

Reach out to us!



Paper: <https://www.usenix.org/conference/usenixsecurity25/presentation/shahid>

Code and artifacts are available online: <https://zenodo.org/records/15693920>



Intelligent • Connected • Secure • Mobile • Systems

References

1. Speaker anonymisation using the mcadams coefficient.
2. Voicemask: Anonymize and sanitize voice input on mobile devices.
3. V-cloak: Intelligibility-, naturalness-& timbre-preserving real-time voice anonymization.
4. Smack: Semantically meaningful adversarial audio attack.
5. Stop bugging me! evading modern-day wiretapping using adversarial perturbations.
6. Micpro: Microphone-based voice privacy protection.