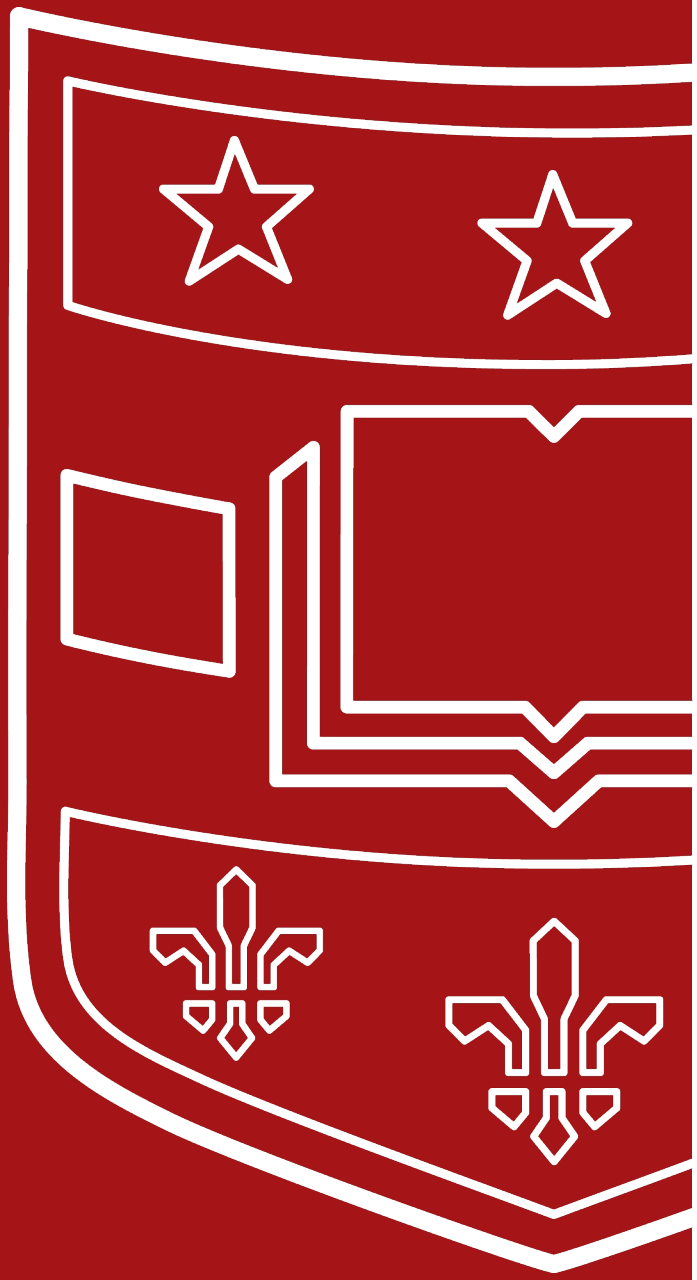
# Security and Privacy in Healthcare

*Ning Zhang*

Computer Security and Privacy Laboratory

Washington University in St. Louis



# What is cybersecurity?




- Cybersecurity is the ***protection of digital assets*** such as devices, networks, and sensitive information from *unauthorized* access, theft, or damage. It involves using technology, policies, and procedures to ***prevent cyber attacks*** and ensure the ***confidentiality, integrity, and availability*** of digital assets.

# Security is a negative science




- A vulnerability is a flaw that is accessible to an adversary
  - Software bugs
  - Design loophole
  - Misconfiguration

***A secure system has to be free of vulnerabilities.***

***How do we know the unknown?***
***A vulnerability is design loop hole that nobody knows about?***

# Inside the mind of hackers



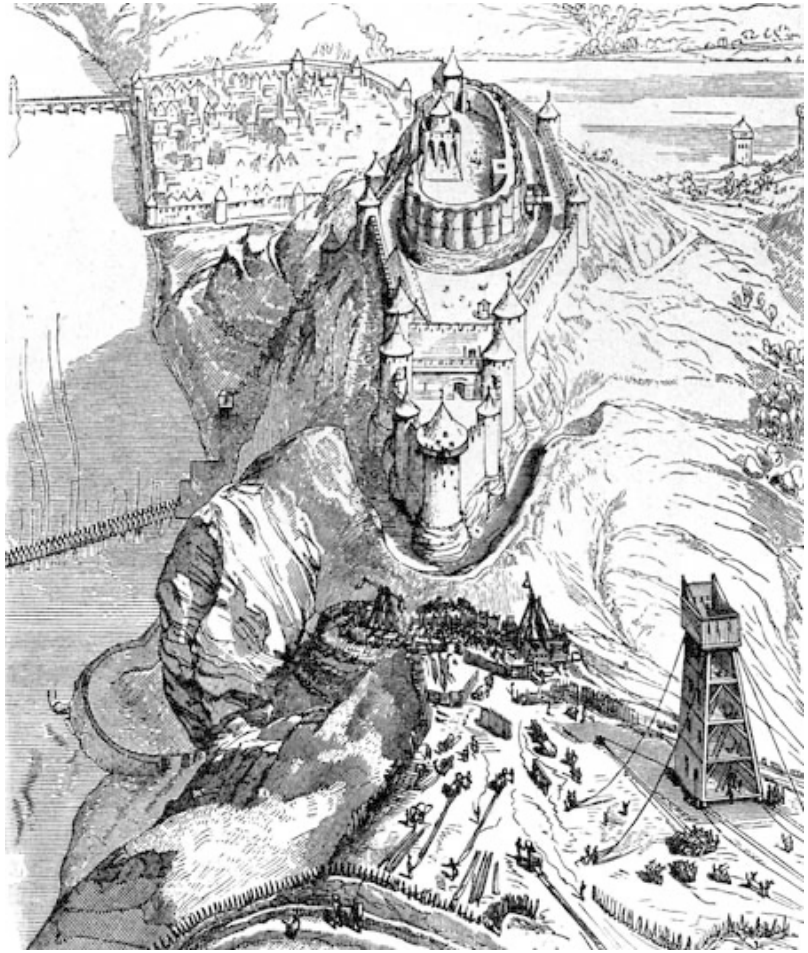

- Everything is hackable, just depends on where you look

- Our methods
  - Look for weakest links
  - Identify assumptions that security depends on
  - Think outside the box

- Explore all possibilities on the attack surface

# Siege of Chateau Gaillard

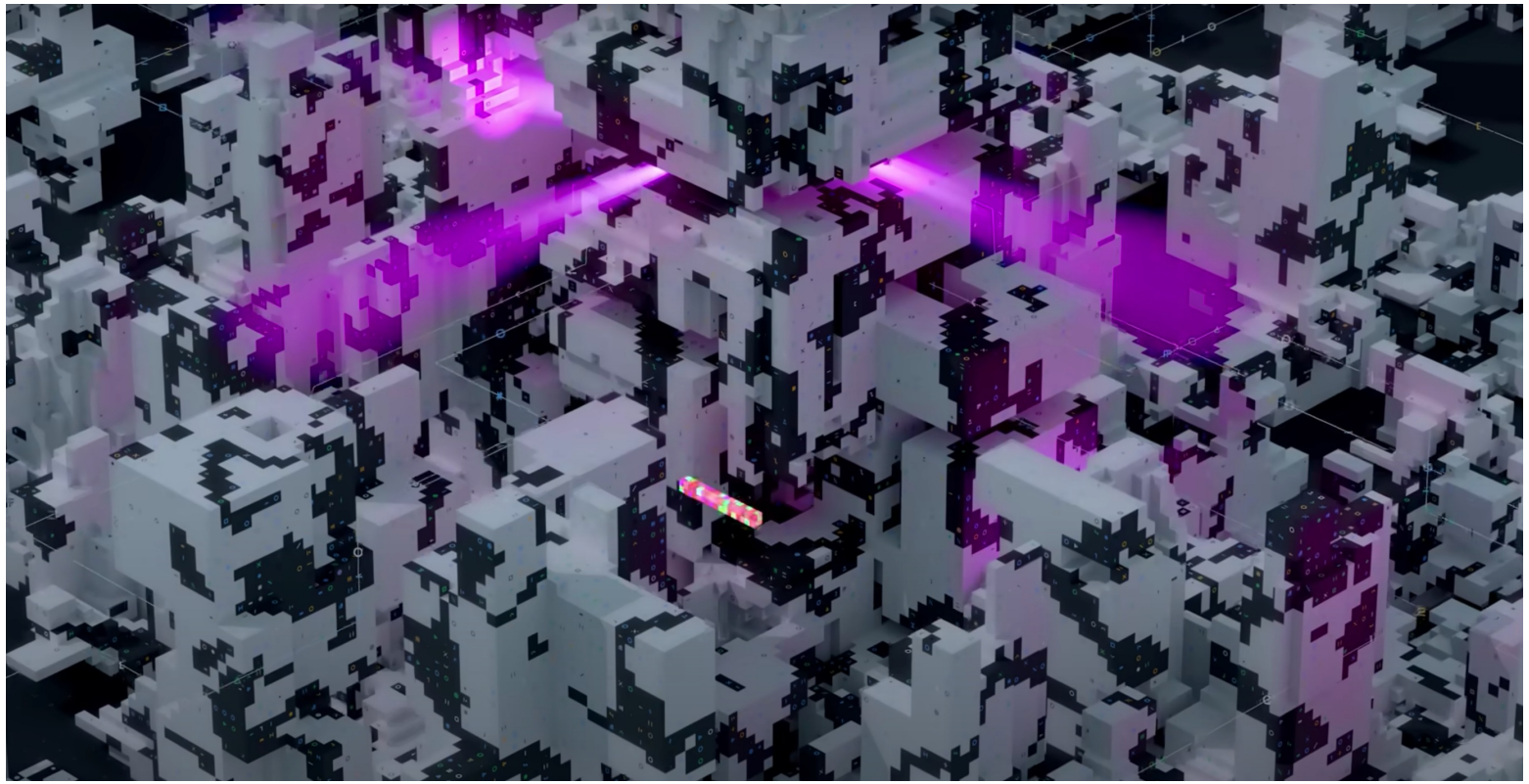
CSPL





*Modern hackers look for such overlooked architectural vulnerability within the castle of computer systems*

CSPL



*Img src: EP005: Project Zero | HACKING GOOGLE*

# Let’s do a small mental exercise



# Congratulations, you are now a hacker!



# A principled approach to secure system



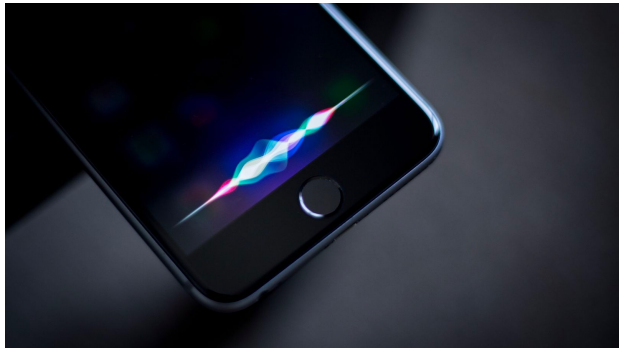

- Threat Model:
  - Your assumptions on the attacker's capability

- Security Goals:
  - Confidentiality
    - Keep the data hidden from unauthorized entity
  - Integrity
    - Prevents unauthorized change to system/data
  - Availability
    - System/data is available when needed

- Security Mechanisms:
  - Encryption
  - Virus scanning
  - Sandboxing

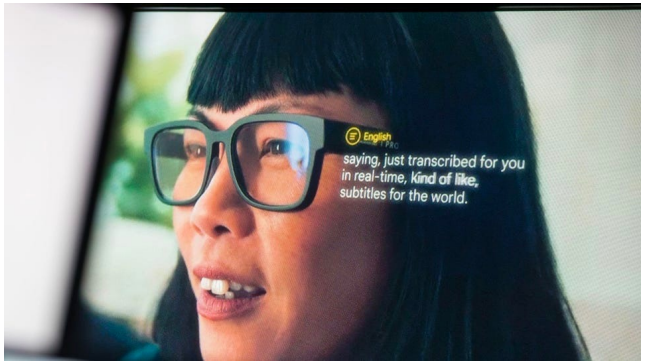
# Life-changing Cyber-physical Systems

CSPL



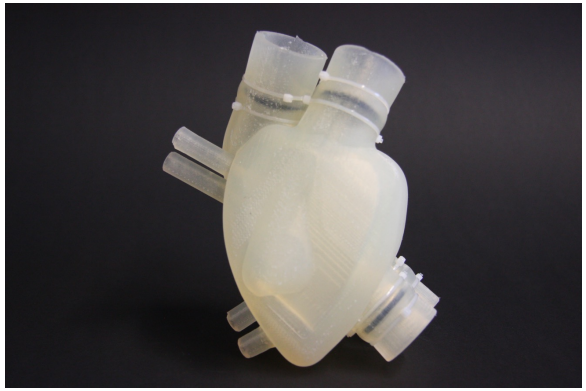
## *Pipeline Attack Yields Urgent Lessons About U.S. Cybersecurity*

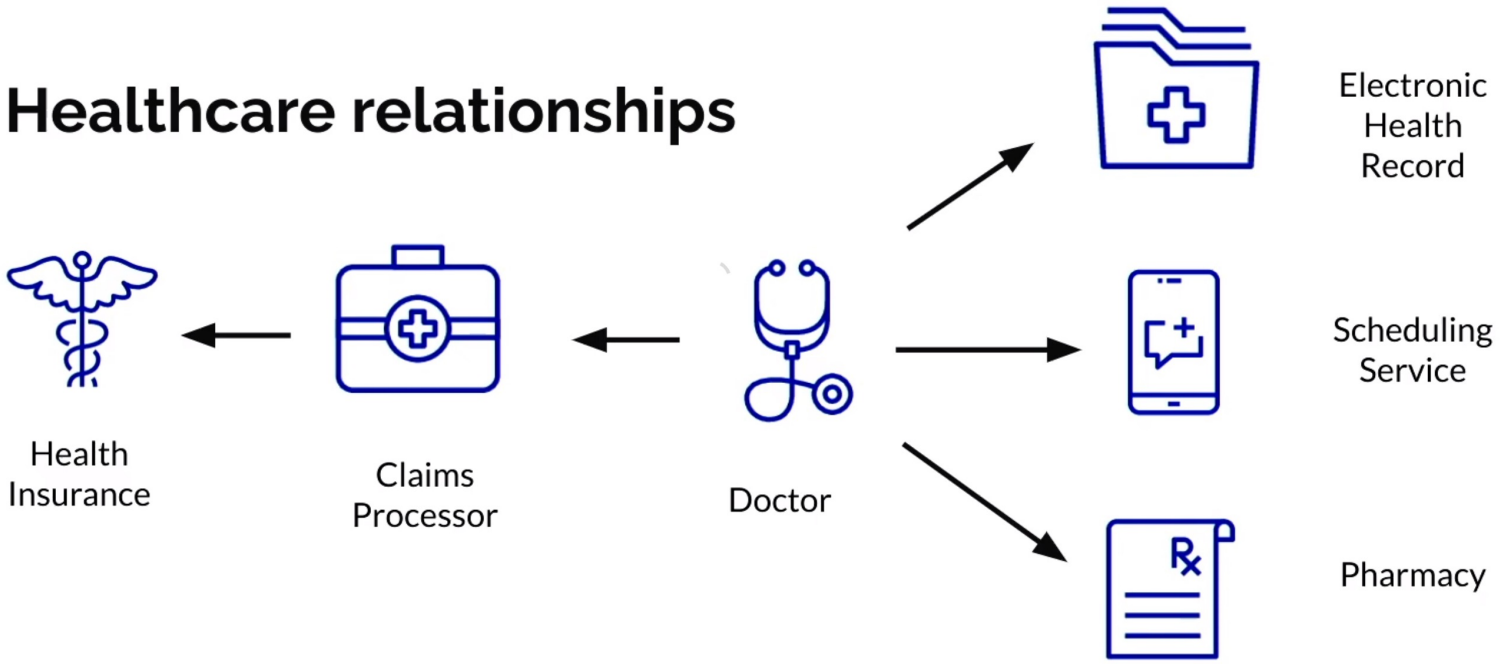
The hack underscored how vulnerable government and industry are to even basic assaults on computer networks.




## Hackers Remotely Kill a Jeep on the Highway—With Me in It

**I was driving 70 mph on the edge of downtown St. Louis when the exploit began to take hold.**




Healthcare relationships
Electronic Health Record
Scheduling Service
Pharmacy
Doctor
Claims Processor
Health Insurance

# Unique Problems in healthcare




- Doctors
  - Focusing on clinical outcome
- Patients
  - Different tech skills
- Nurses and staff
- Regulators
  - Do they understand?
- Insurance
  - Financially driven
- Tech companies
  - Wearables
  - Cloud



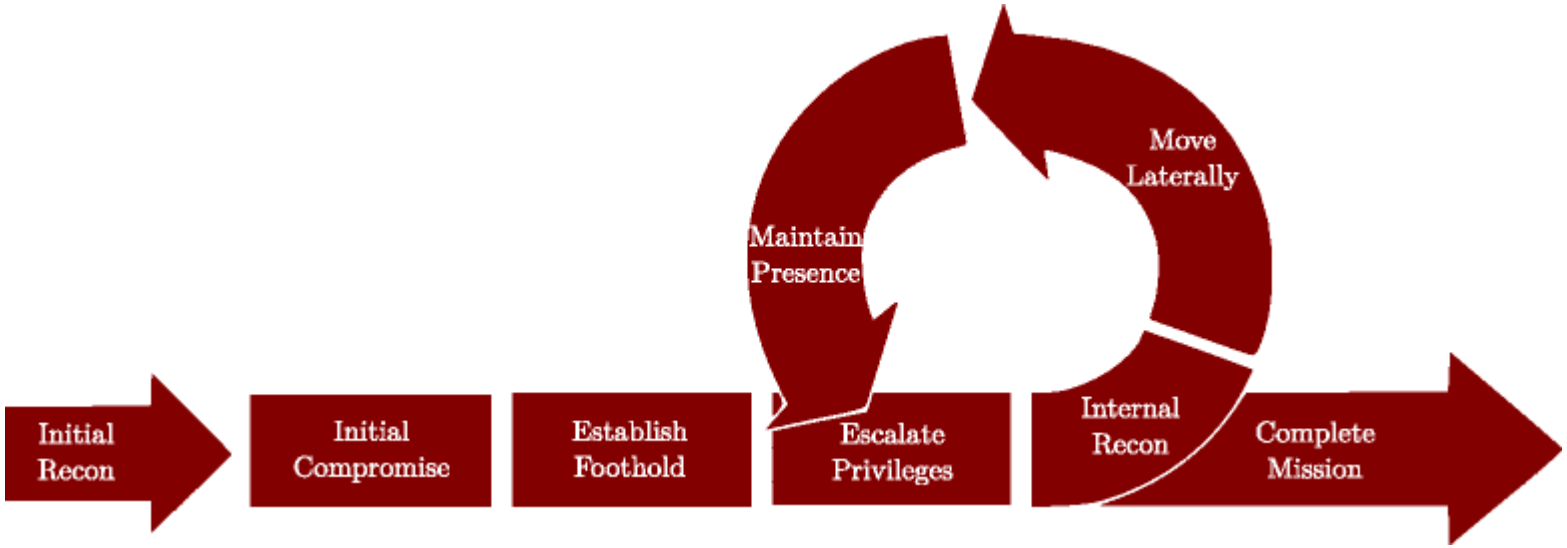

# System Protection for Medical Devices

# CPS/IoT system architecture

CSPL



# Dissecting the internals of a hack



# Hacking Process - Reconnaissance

CSPL

***Gathering information of the target system***



Network Scanning and Mapping

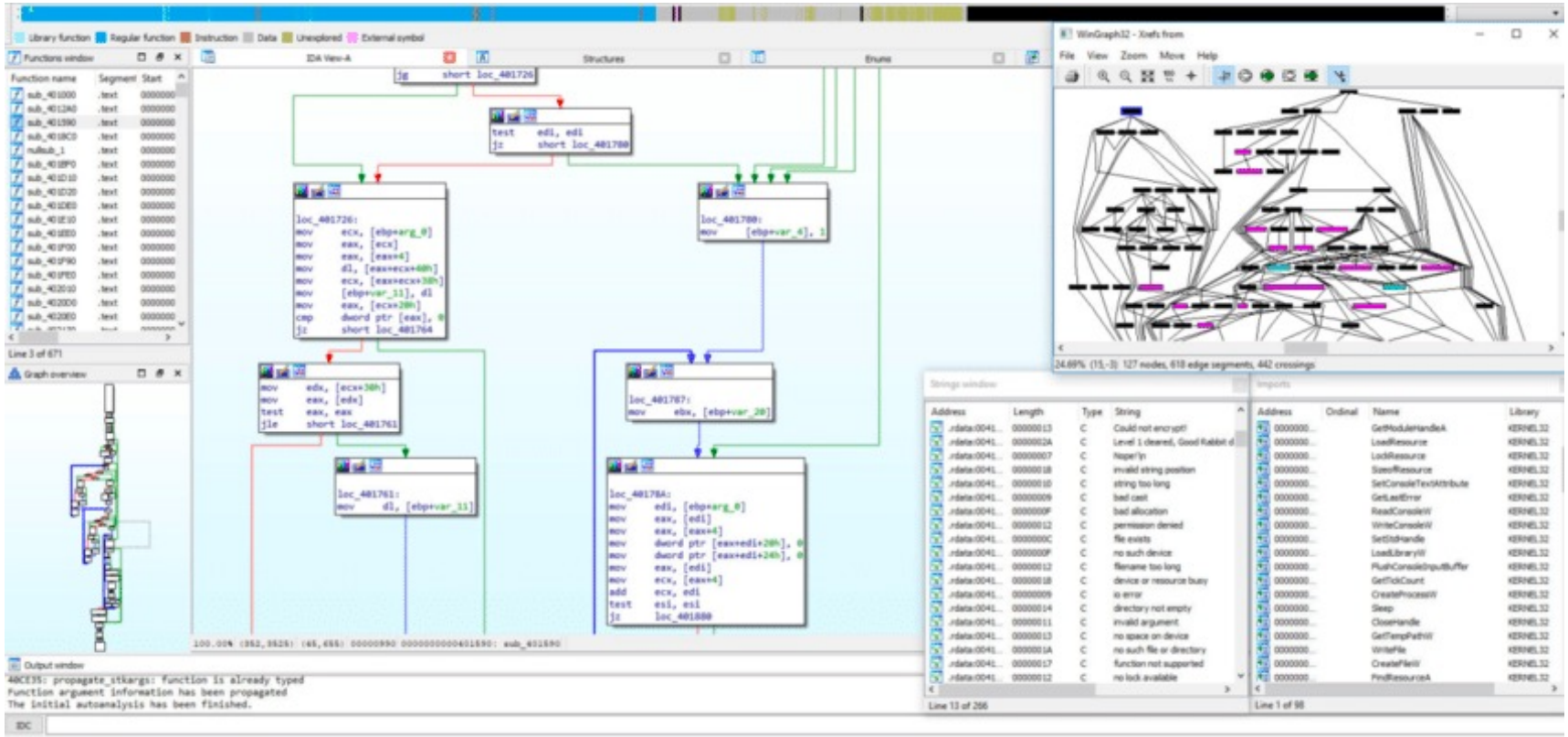


Understanding its internals

# Hacking Process – Vulnerability Identification

CSPL

***Analyze the target system to identify the vulnerabilities***

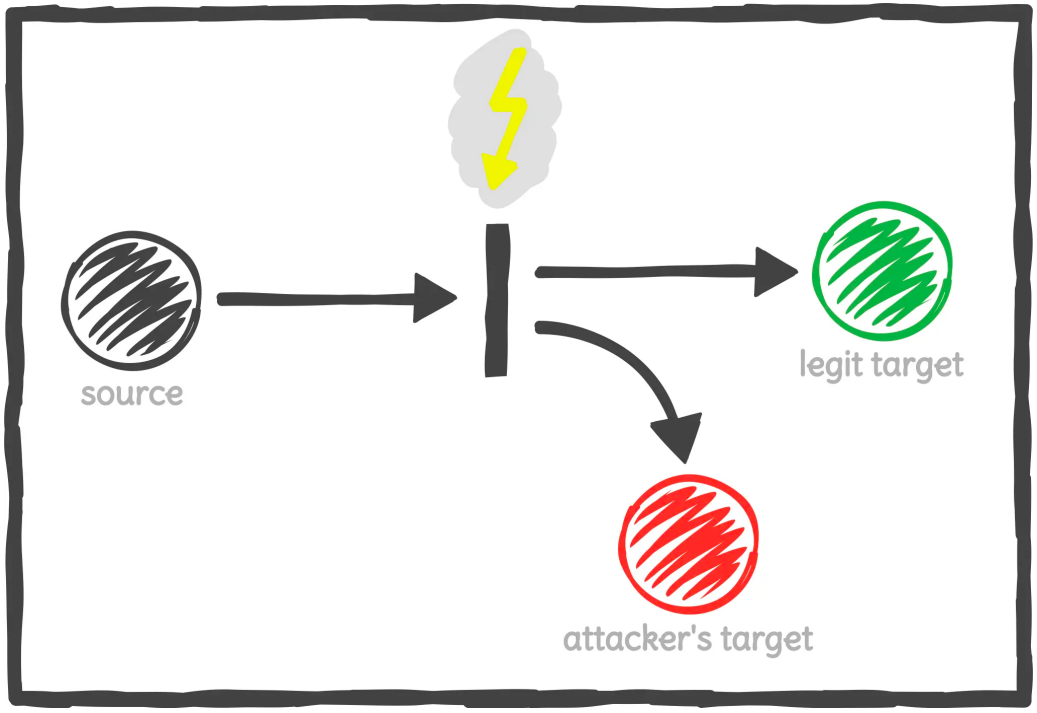


# Hacking Process – Working principle behind exploitation



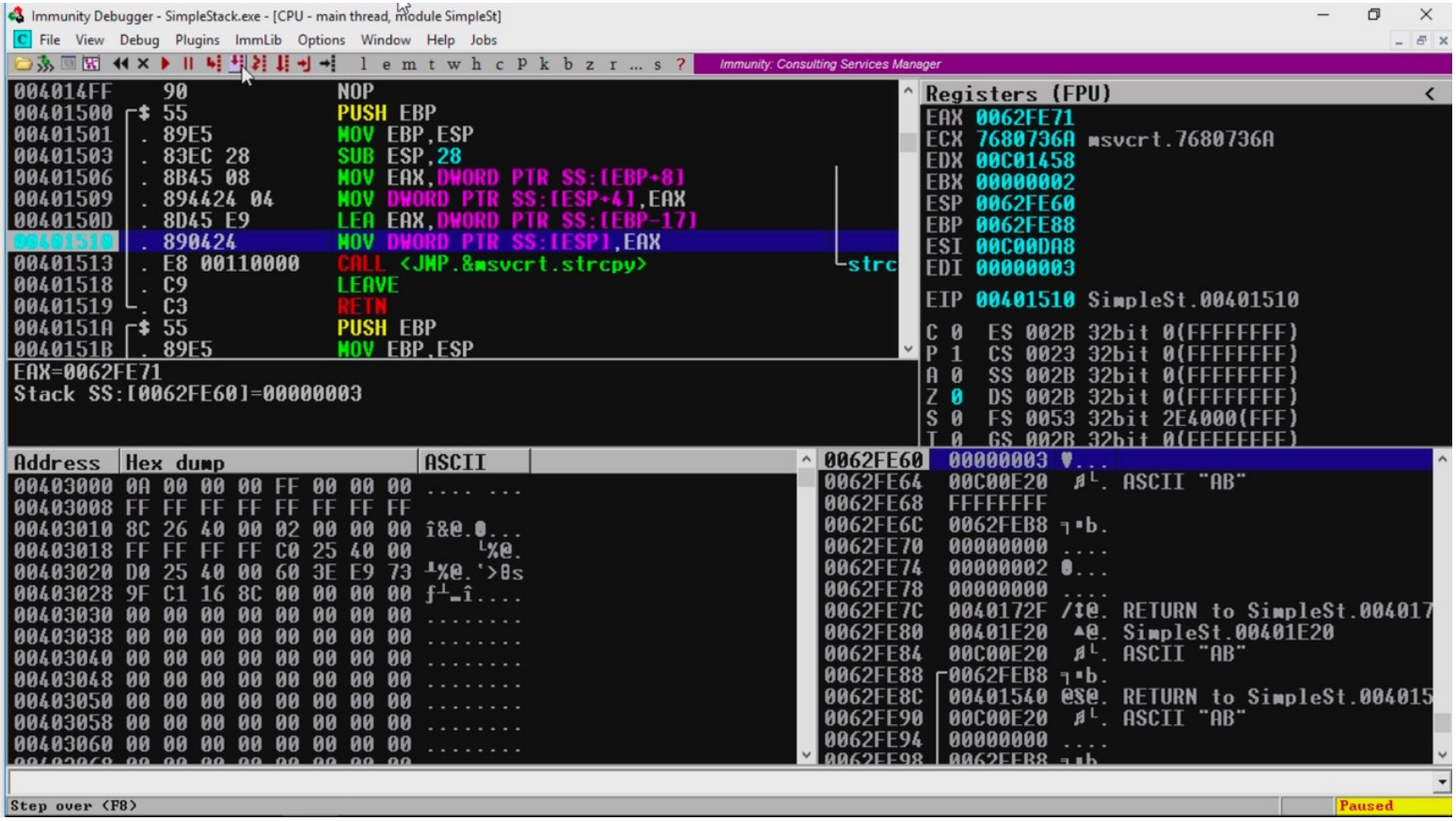

If (a > 10) call function A

Else call function B



# Hacking Process – Exploit Development

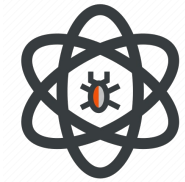
CSPL



# Exploit Deployment

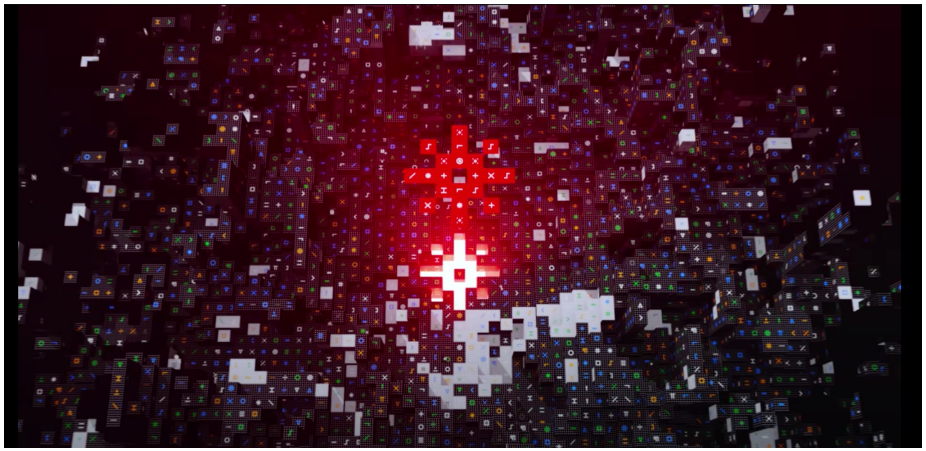
CSPL



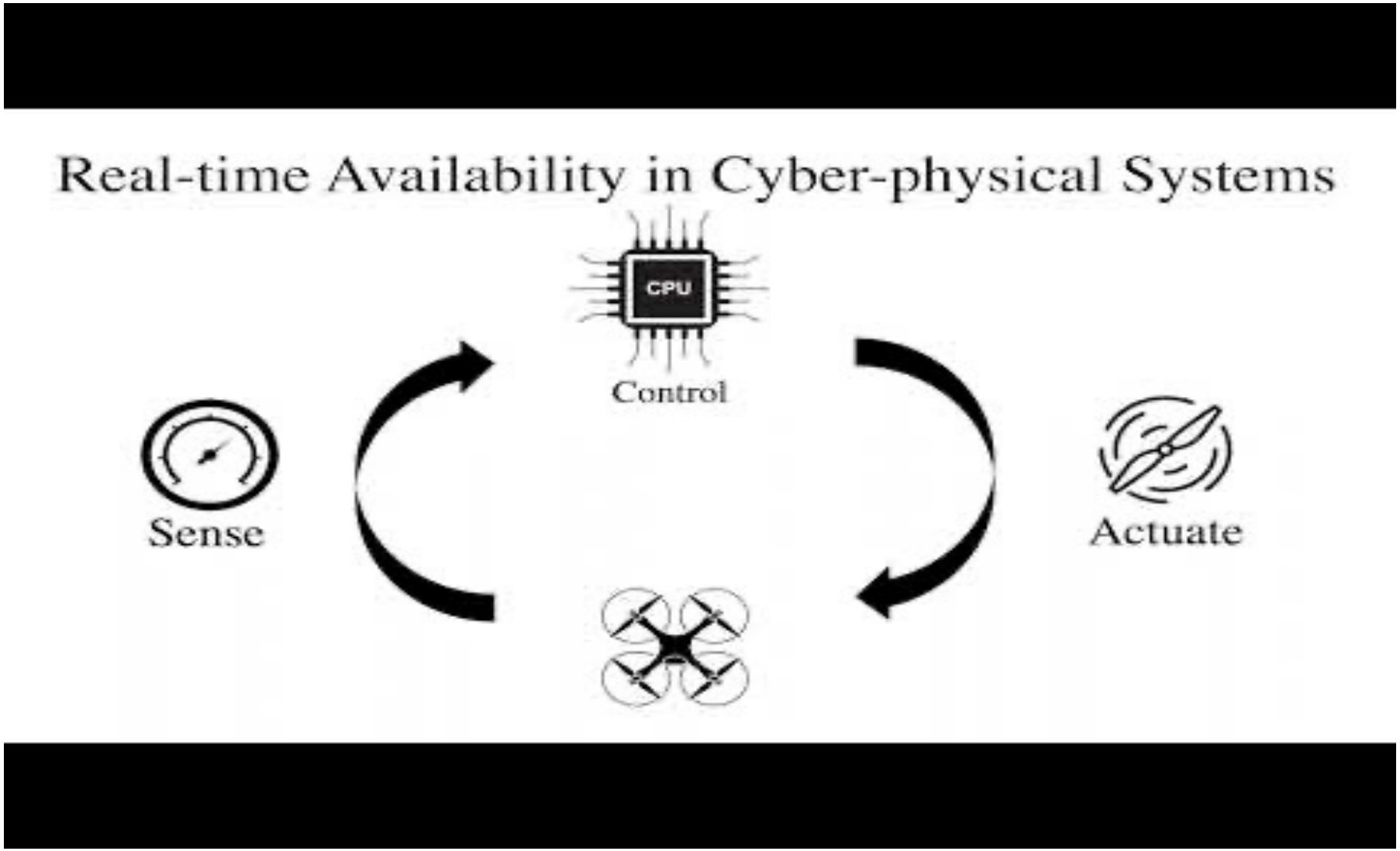






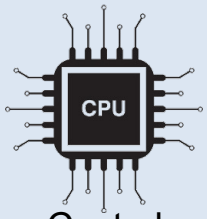


# *RT-TEE: Real-time System Availability for Cyber-physical Systems using ARM TrustZone*



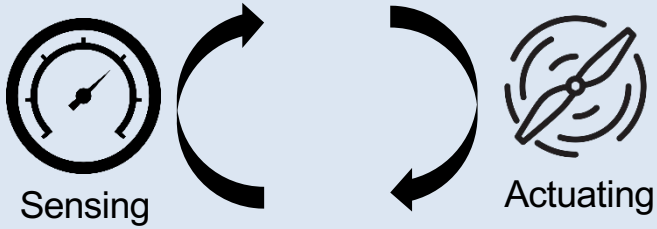



# Real-time Availability in Cyber-physical Systems




Control

***Real-time Computation Availability ( Control )***




***Real-time I/O Availability (Sensing & Actuation)***



Cyber-physical System




# Medical AI Security

# A correct image classification



99.99% it is Golden Retriever

*Img src: Nicholas Carlini @ Google Brain*

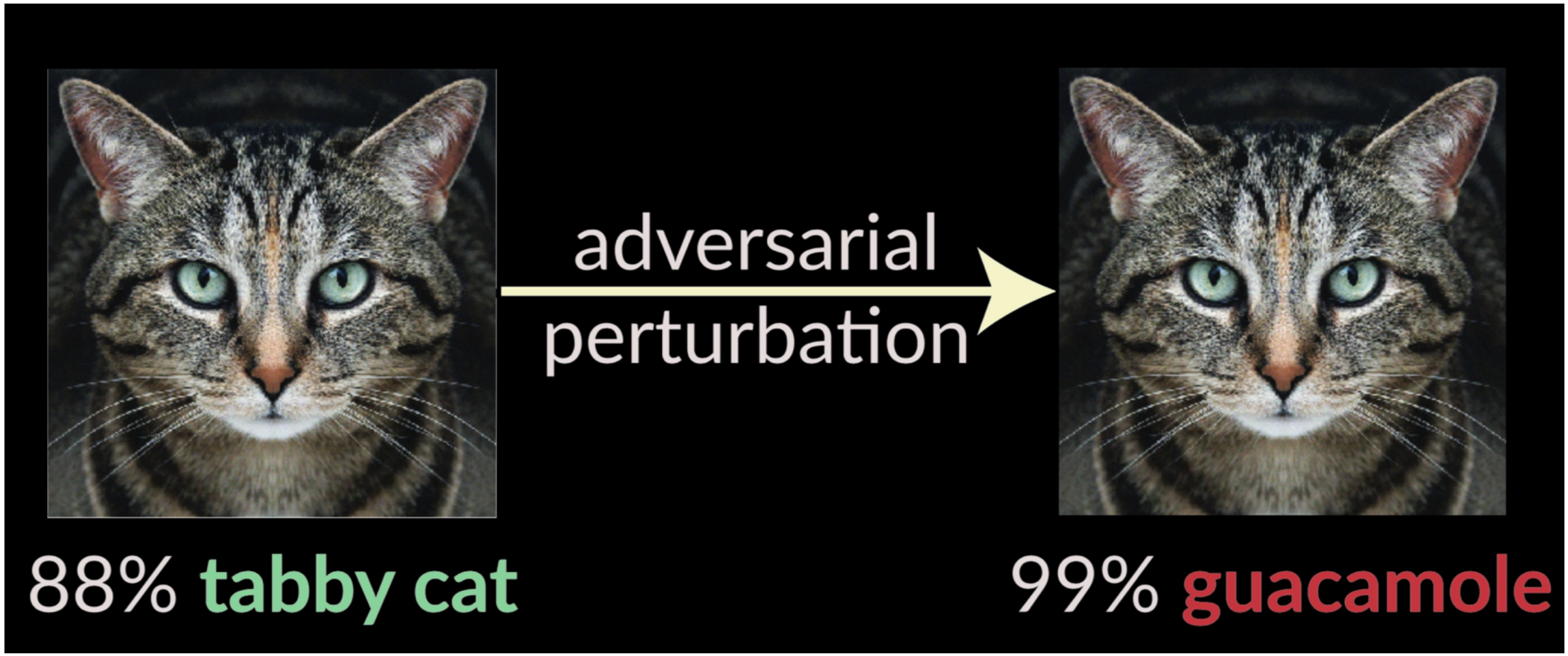
# A "misguided" image classification



99.99% it is Guacamole

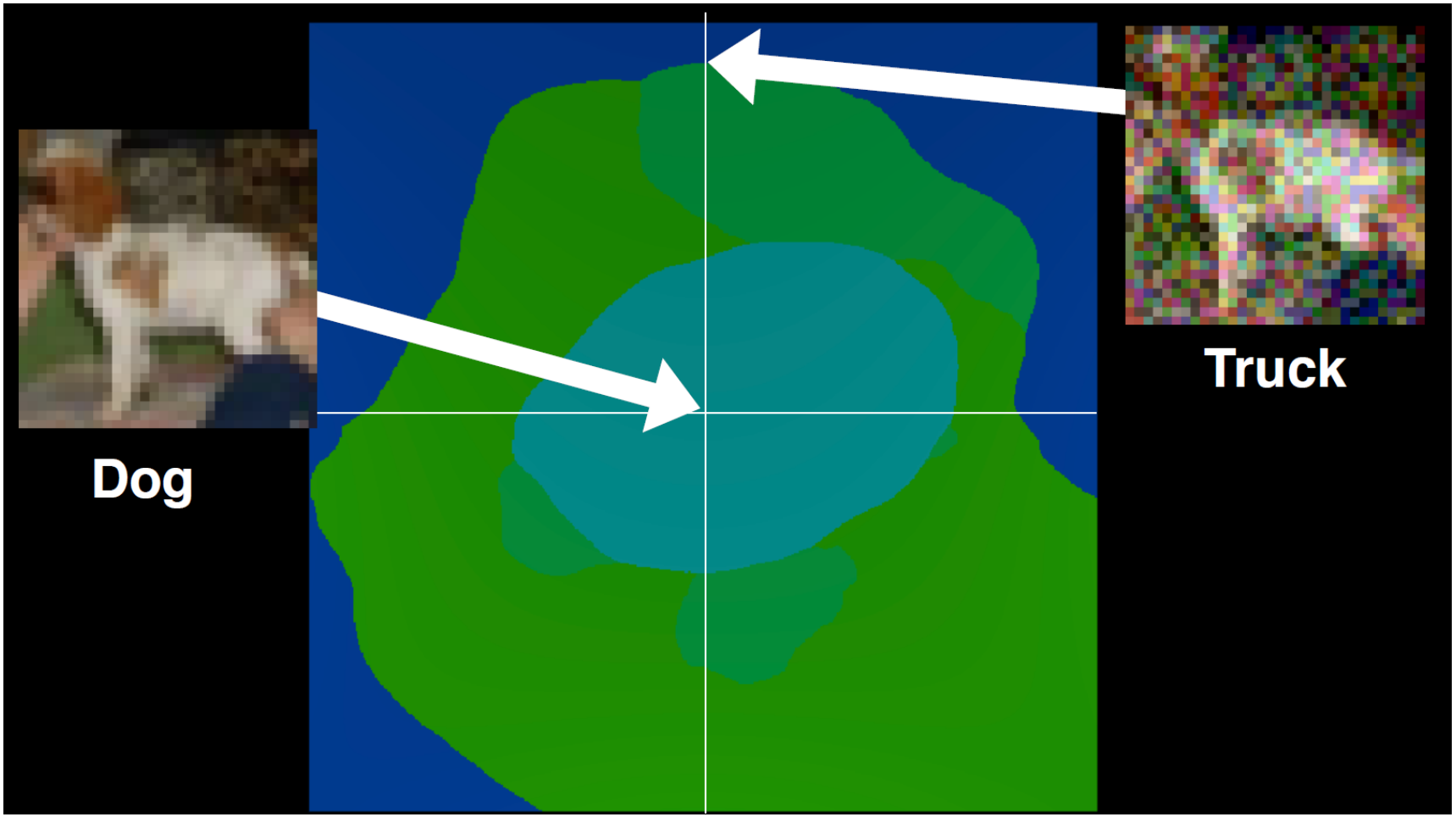
*Img src: Nicholas Carlini @ Google Brain*

# Using adversarial perturbation, we can change the image classification



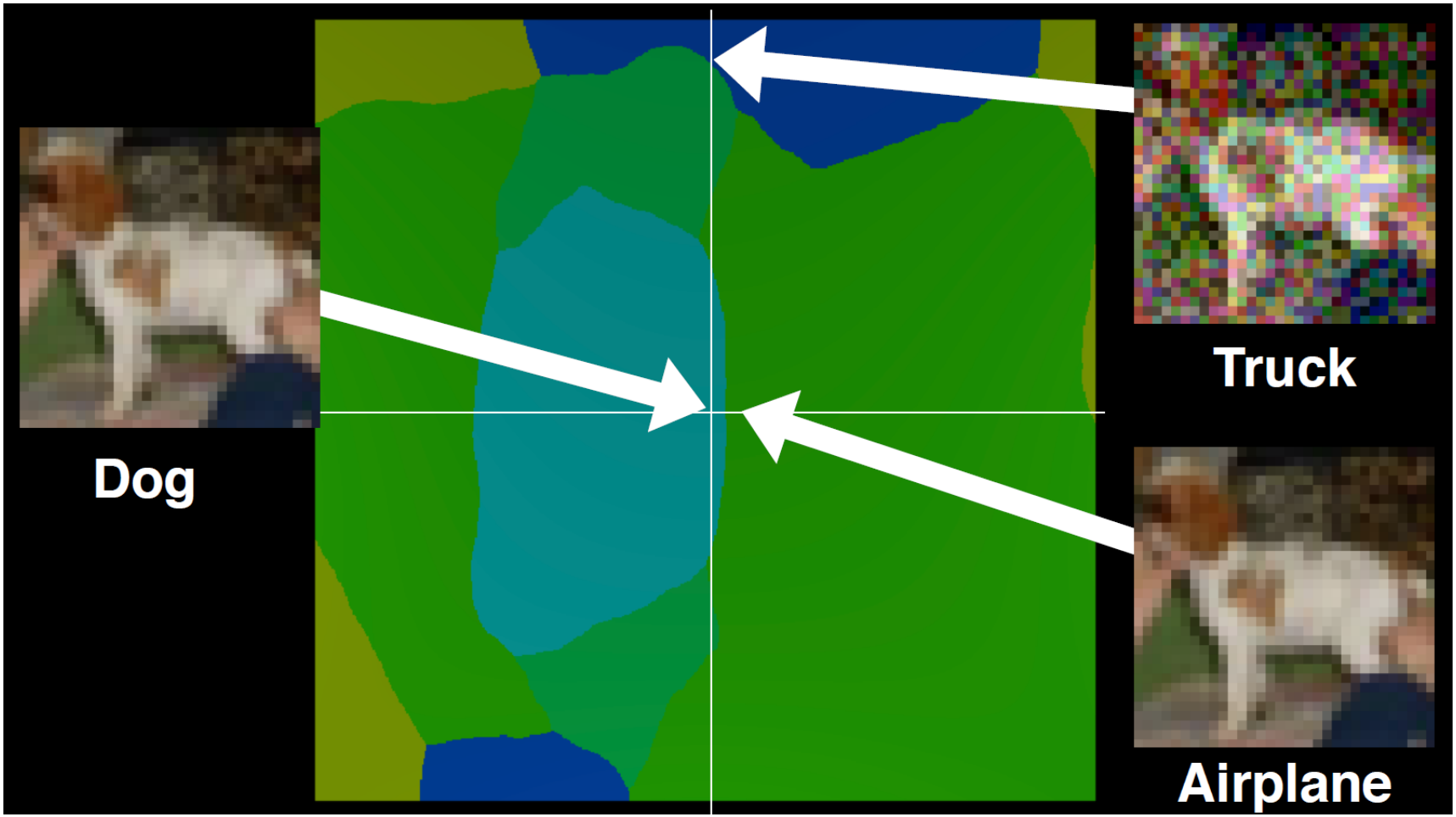

*Img src: Nicholas Carlini @ Google Brain*

# Areas that effectively classify the image




*Img src: Nicholas Carlini @ Google Brain*

# Areas that effectively classify the image



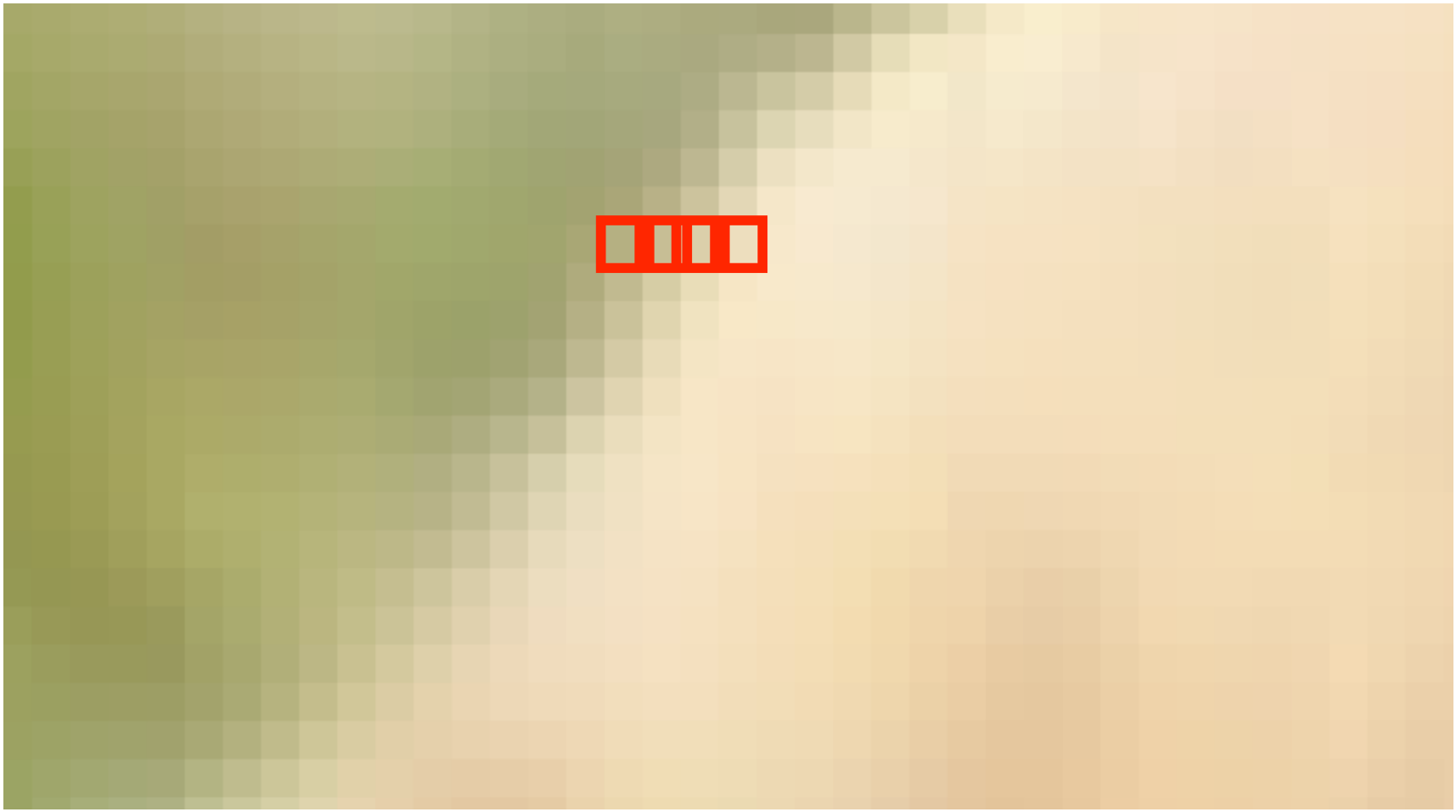

*Img src: Nicholas Carlini @ Google Brain*

# Looking at the adversarial example again



*Img src: Nicholas Carlini @ Google Brain*

# The miniscule yet effective perturbation



*Img src: Nicholas Carlini @ Google Brain*

# We can even achieve this...

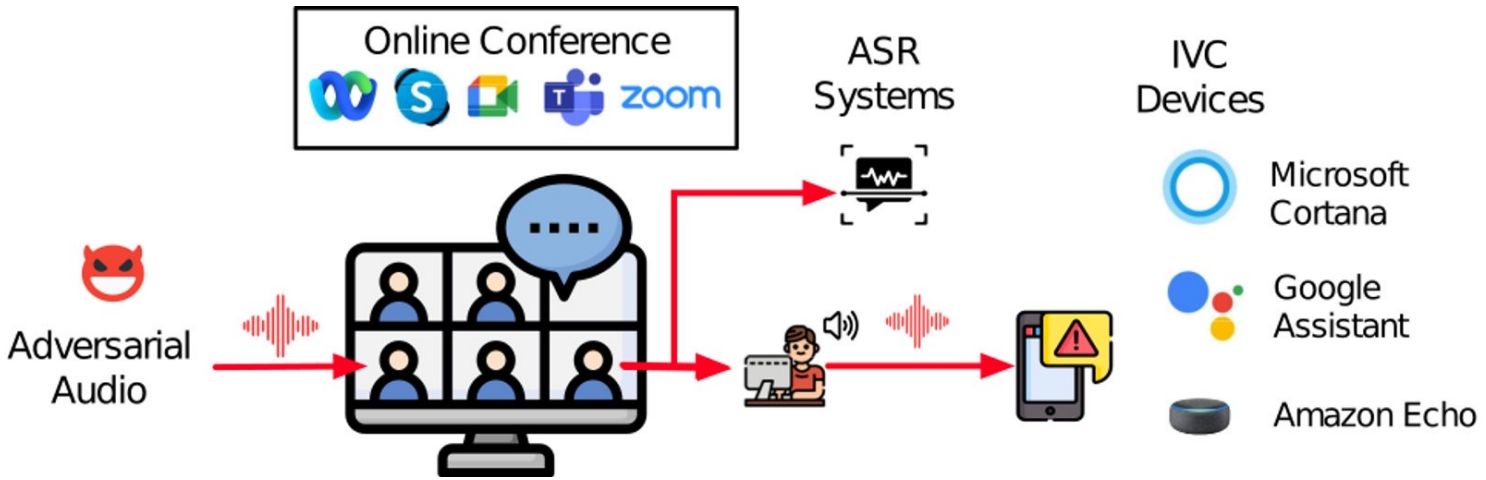



*Img src: Nicholas Carlini @ Google Brain*

# Breaking AI System - Evil Calls

CSPL

- Deliver adversarial audio over online conference
- Propose a query-based method to craft adversarial audio tailored to VoIP




*Han Liu, Zhiyuan Yu, Mingming Zha, XiaoFeng Wang, William Yeoh, Yevgeniy Vorobeychik, Ning Zhang. "When Evil Calls: Targeted Adversarial Voice over IP Network." ACM CCS 2022.*

# Our Work on AI Security – Evil Calls



Transcript: call my wife



Transcript: OK Google



Transcript: call my wife

# Result




- We perform the attack on four popular voice assistants over five video conferencing software
- Over-the-line attacks succeed with a better SNR in only 1500 queries
- We also demonstrate it is possible to perform over-the-air attack without getting highlighted by zoom











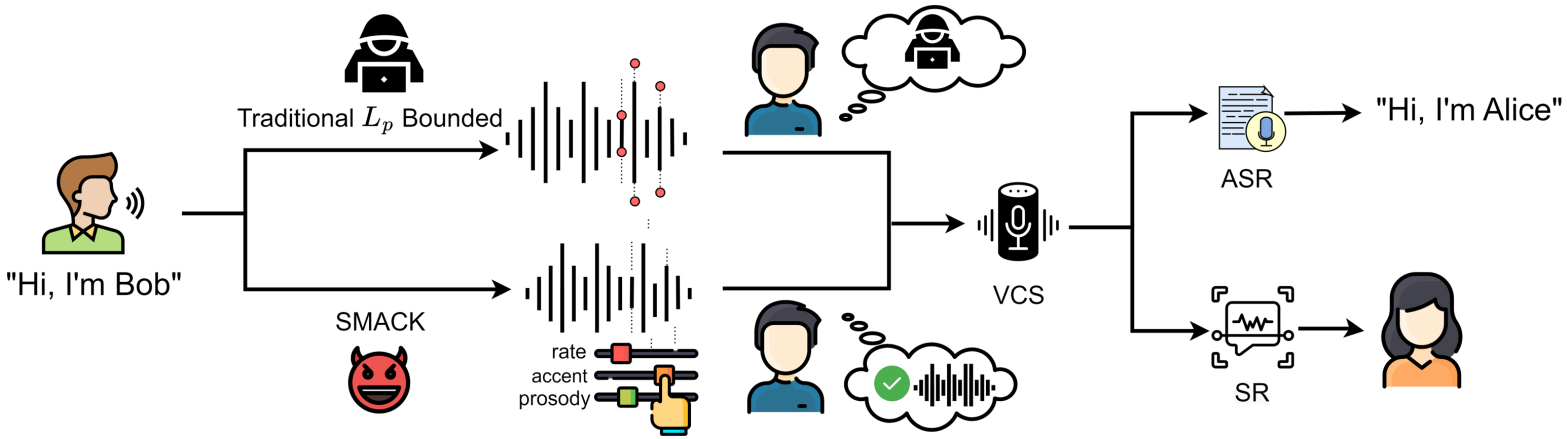





*https://sites.google.com/view/targeted-adversarial-voip*

# How about not using noise at all ?

***How could adversarial attacks against AI be even stealthier?***




Yu, Zhiyuan, Yuanhaur Chang, Ning Zhang, Chaowei Xiao.
"SMACK: Semantically Meaningful Adversarial Audio Attack", *USENIX Security Symposium 2023*

# Adversarial Examples



To human: "Wipe the floor"

To Voice assistant: "Open the door"



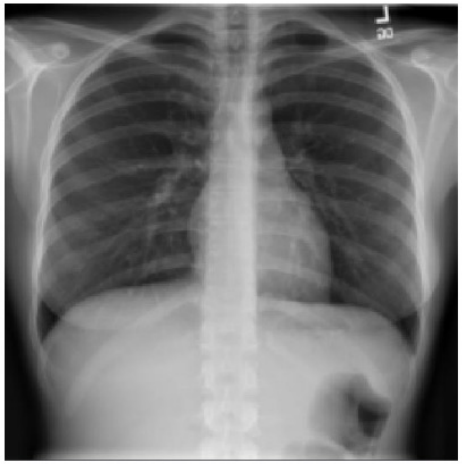
To human: "Wipe the floor"

To Voice assistant: "I want to share"

# Applied in healthcare

CSPL

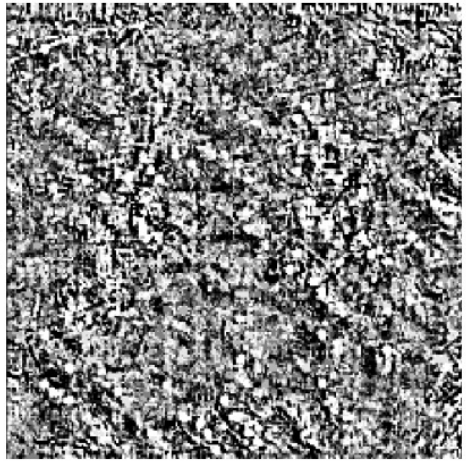
**Clean Scan**



**Prediction**: Healthy
**Confidence**: 99.6%

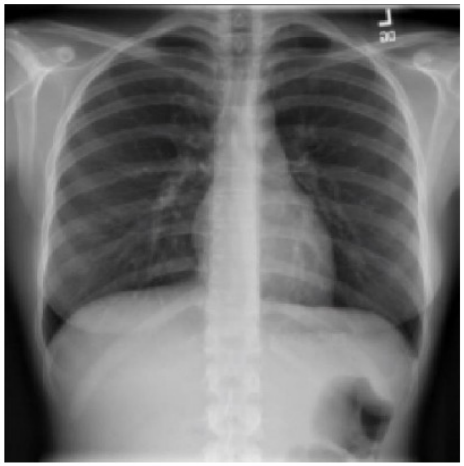
+ 0.01x

**Adversarial Noise**



=

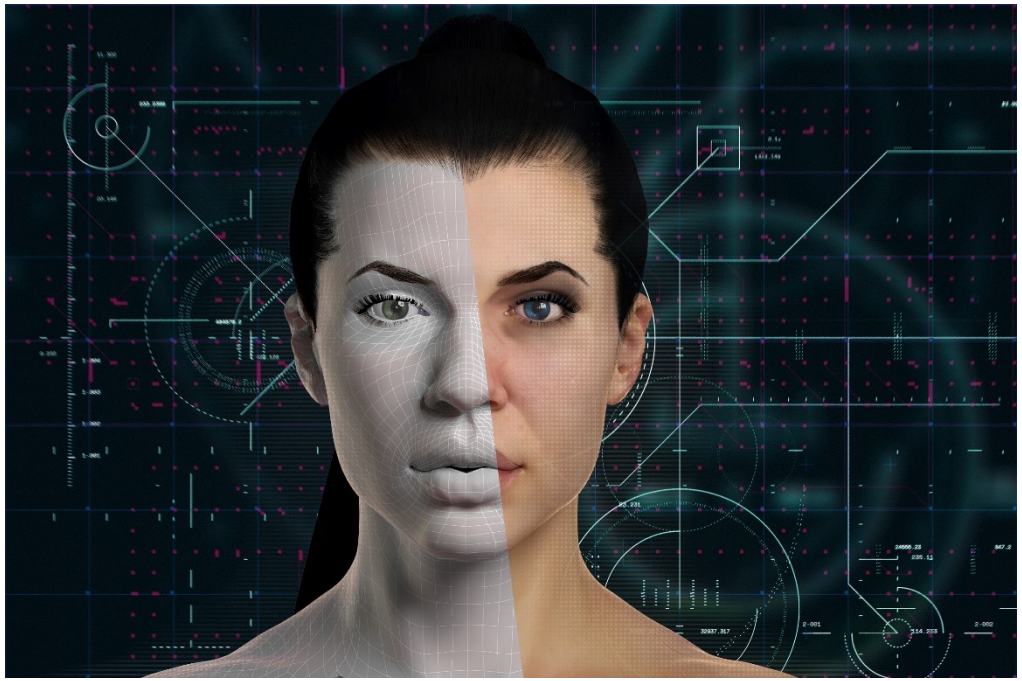
**Adversarial Scan**



**Prediction**: Pneumonia
**Confidence**: 99.9%

CSPL

# Generative AI and Misinformation



# 2.6 Million AI-generated Faces

CSPL



*Img src: https://www.forbes.com/sites/johnkoetsier/2021/04/11/26-million-fake-faces-ai-generated-people-are-the-future-of-the-metaverse-celebrity-and-perhaps--all-of-us*

# Seeing is believing?

CSPL



# List of the generative AI companies here




OpenAI

SYNTHESYS

GLIACLOUD

invideo

synthesia

DEEPBRAIN AI

Fliki

elai

Rephrase.ai

runway

lumen5

# Deep fake can be misused!




## Emma Watson reads Mein Kampf while Biden announces invasion of Russia in latest AI voice clone abuse

Updated on: 01 February 2023

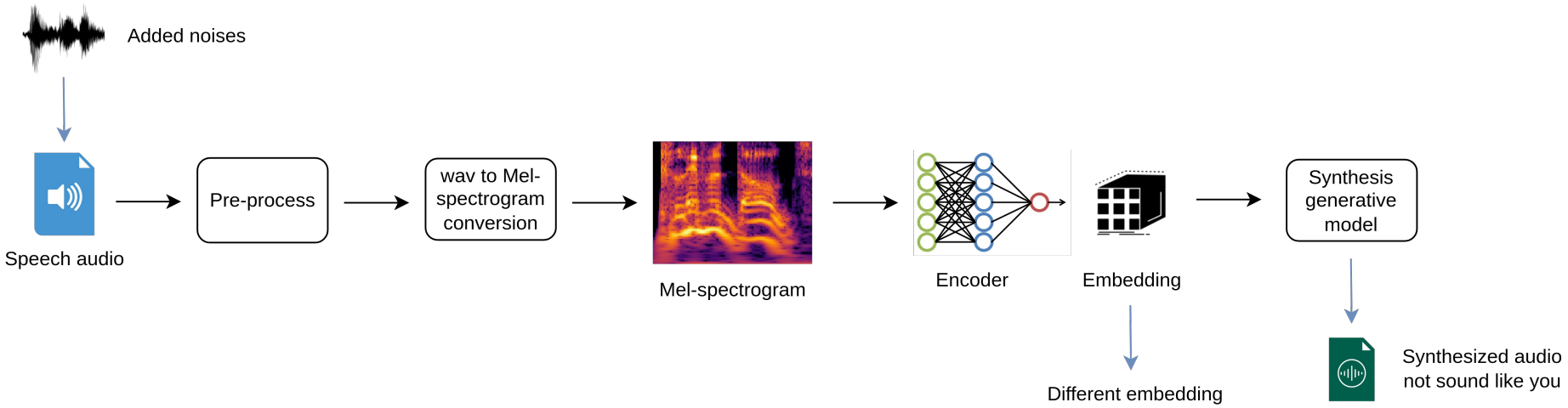
ARTIFICIAL INTELLIGENCE / TECH / CREATORS

## 4chan users embrace AI voice clone tool to generate celebrity hatespeech

# AI Approach: Using adversarial technology for good - Preventing unauthorized Deepfake

CSPL

- DeFake protects user voice by making unauthorized synthesis fail!




- Demo

Original synthesized audio

Synthesized audio with perturbed speech

# Will you trust AI?

CSPL









# Privacy vs XAI

# Privacy Risks of Explainability of XAI




- Explainability remains a fundamental challenge in the field of deep learning.
- XAI has been employed in various critical domains, such as medical diagnosis and computer security, due to its explainability into the model decision-process
- Little attention has been given to the privacy risks of XAI, especially considering the sensitive nature of the data in its applications

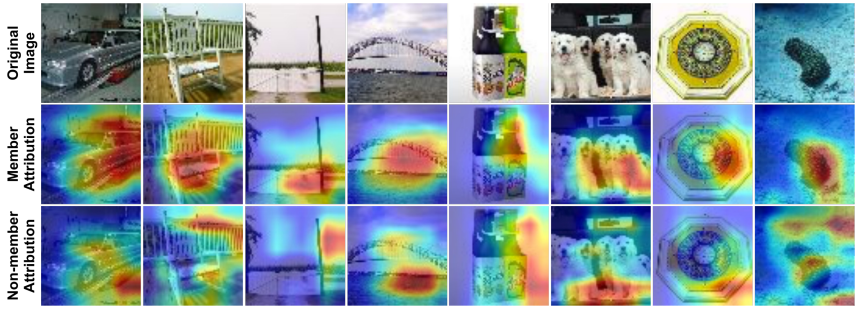
Han Liu, Yuhao Wu, Zhiyuan Yu, and Ning Zhang.
"Please Tell Me More: Privacy Impact of Explainability through the Lens of Membership Inference Attack",
IEEE Symposium on Security and Privacy (Oakland), 2024.

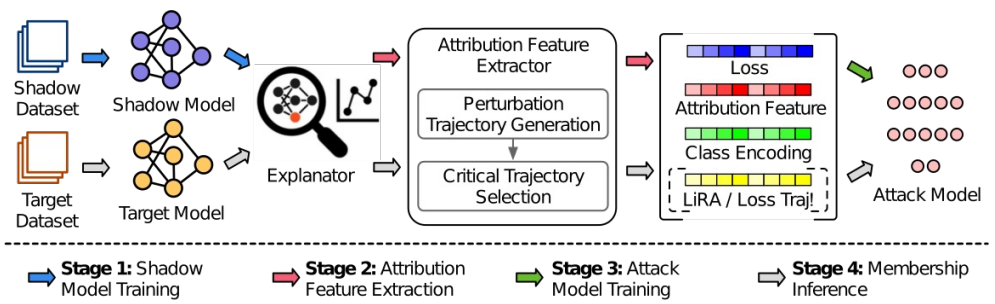
# Design Insights and Methods

CSPL

- The attribution maps of explanators focus on informative features for member samples than non-member samples

- The adversary could perturb images and distinguish between members and non-members using prediction changes under the guidance of explanators






# Emerging Technologies

# 3D Printing Is Bringing Revolutionary Changes to Healthcare – Saving Lives

CSPL

Posted in | News | 3D Printing

**How Seattle Children's Hospital is using 3D-printing to Save Lives**

**Doctors Print Up '4-D' Airway, Save Three Babies**

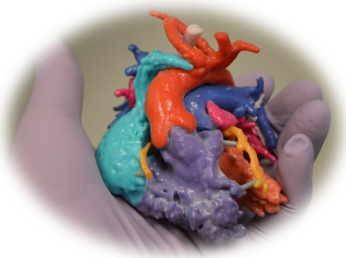
Three small children with weak airways that were constantly collapsing are alive and well thanks to a "4-D" printed splint that saved them from almost certain suffocation.

SPONSORED POST

3D printers have proved their value to hospitals for limb-saving surgery, Covid-19 supply chain issues, and more

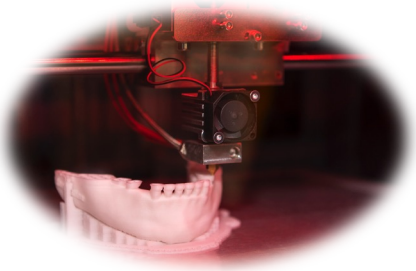
**3-D Printed Medical Device Saves a Life for the First Time**

3-D printers are increasingly being used in medical settings.

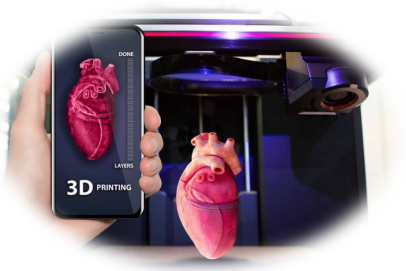






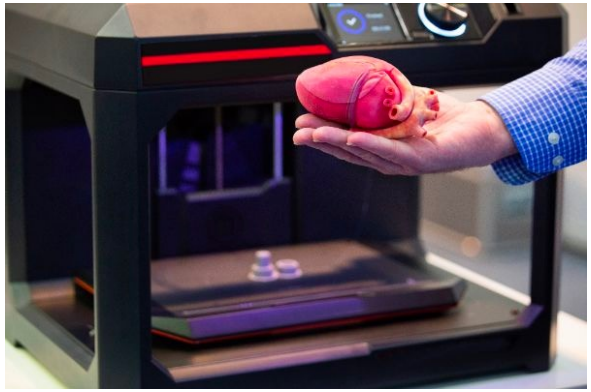






# Attacks on these systems can lead to life-death for patients

CSPL

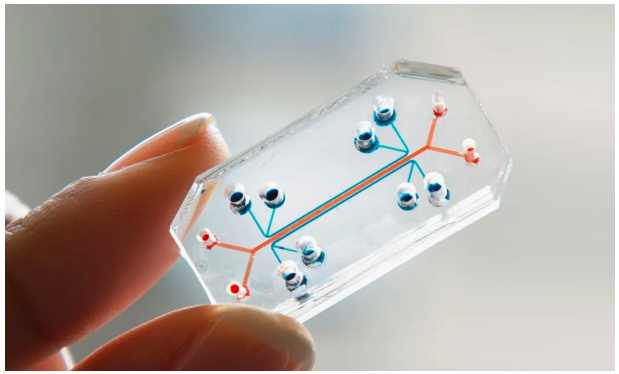


Tissue Engineering



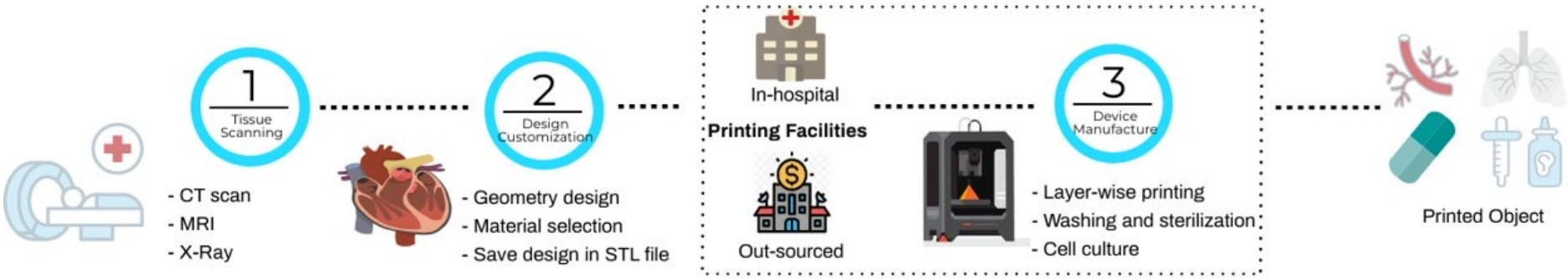
Implants

Organs-on-chips



# How Medical Printing Works?

CSPL



# Threats in the Printing Pipeline

CSPL

- Existing work has demonstrated cyber-physical threats in various forms



Materials



Controller PC

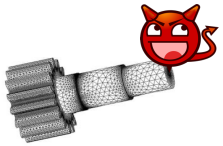


Firmware



Manufacture parameters



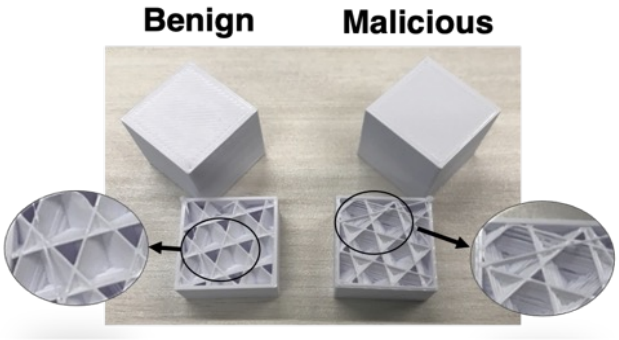


STL file

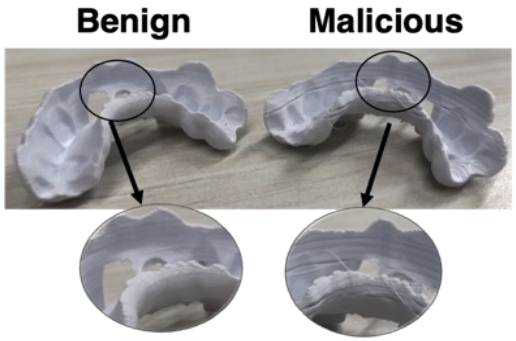
# Our PoC Attack on Medical Printing




- Misuse temperature control to cause malfunction of material solidation
  - Target: Fused deposition modeling (FDM)
  - Stealthy: Does not change nozzle trajectory (commonly monitored for defense)




(a) Attacked cube model




(b) Attacked dental guide

# Existing Efforts in Protecting 3D Printing




- The majority of existing work focuses on "Procedure-based" defensive approaches
  - Vision-based: Digital photos of printing processes for monitoring [1]
  - Side-channel-based: Audio signature; Vibration; Magnet [2][3]

- They showed impressive performance
  - However, some unique challenges present in medical printing...

# Challenges to Protect Medical Printing




- Medical printing technology is diverse
  - FDM, multi-jet fusion (MJF), stereolithography (SLA), liquid-based extrusion, etc.
- Printing process is complicated
  - Difficult for “Procedure-based” methods to monitor every manufacture process
- Medical devices are patient-specific
  - Defenses relying on prior data (e.g., side-channel signature) are less effective
- Results are less interpretable to medical practitioners with less familiarity in security

# Our Approach – XCheck Overview

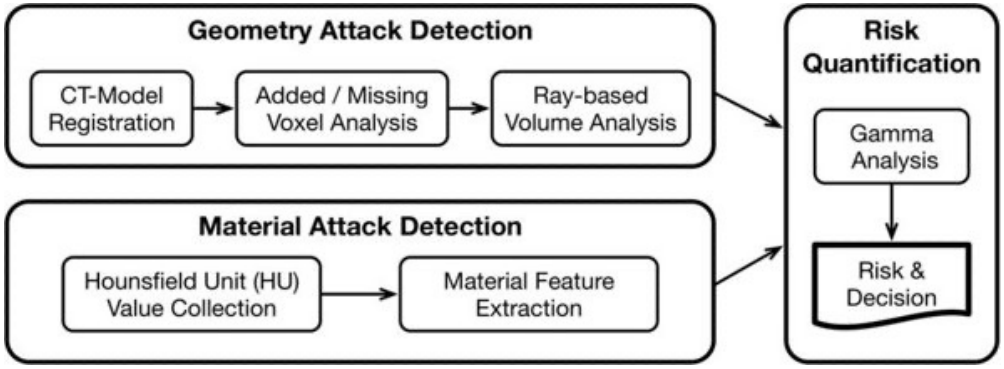



- Key idea: Check the final product against its digital design ("Product-based" defense)
  - Reconstruct from physical object: CT scanning that is naturally accessible to medical printing facilities
  - Validate multiple aspects of prints: Geometry and material
  - Enhance usability with interpretable detection results: Visualizing the potential malicious areas and quantifying risks with a score

# Defense in Depth – XCheck Designs

CSPL

- Key design elements
  - Geometry attack detection: registration & computational graphic techniques
  - Material attack detection: HU value distribution analysis
  - Risk quantification: inspired from medical gamma analysis

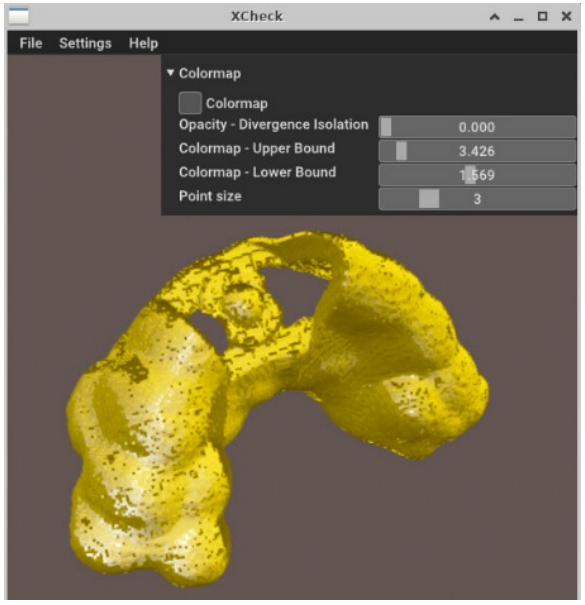


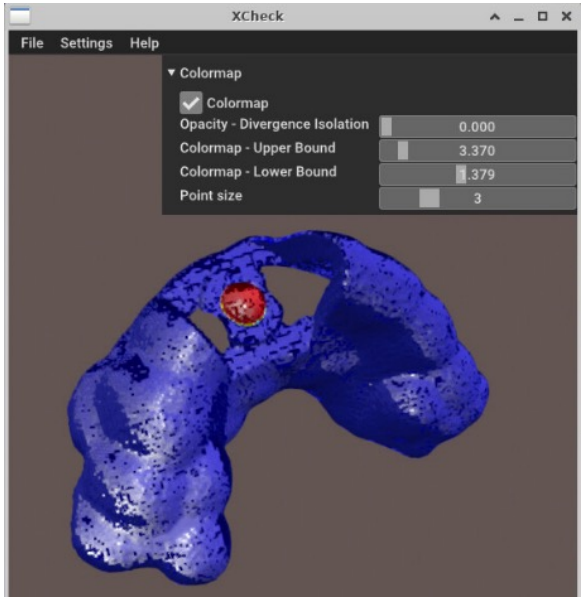
# XCheck – Geometry Verification

CSPL

Our Key idea:

Using CT imaging to reconstruct the digital clone of the printed device, then compare with the original design.



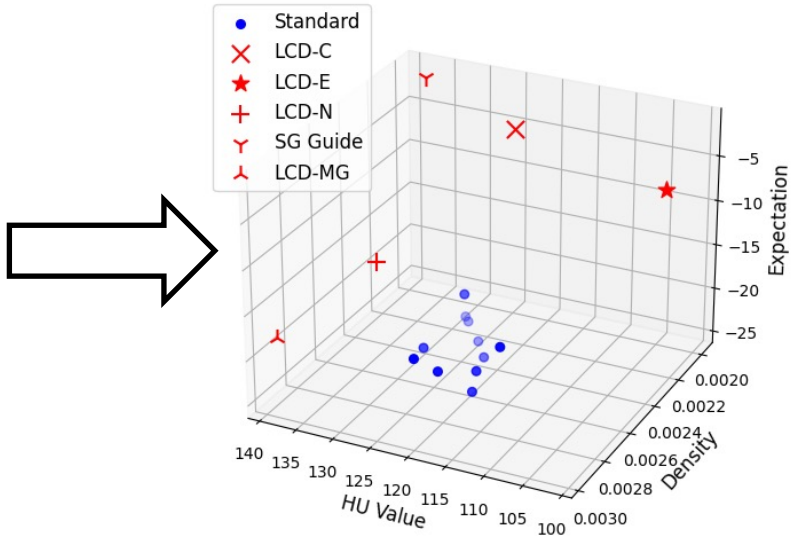
# XCheck – Material Verification

CSPL

- Different materials yield varying attenuation of electromagnetic rays, such as CT value.
- We can characterize material based on this



Experiments with bone screws printed with different materials



# Evaluation – Printed Devices




- By working with medical practitioners in WashU medical school, we...

- We printed and scanned 47 unique PSDs
  - 4 Applications: Bone scaffold, dental guide, orthopedic screw, lung-on-a-chip
  - 7 Materials: Including bio-compatible materials for implants

# Evaluation – Results

CSPL

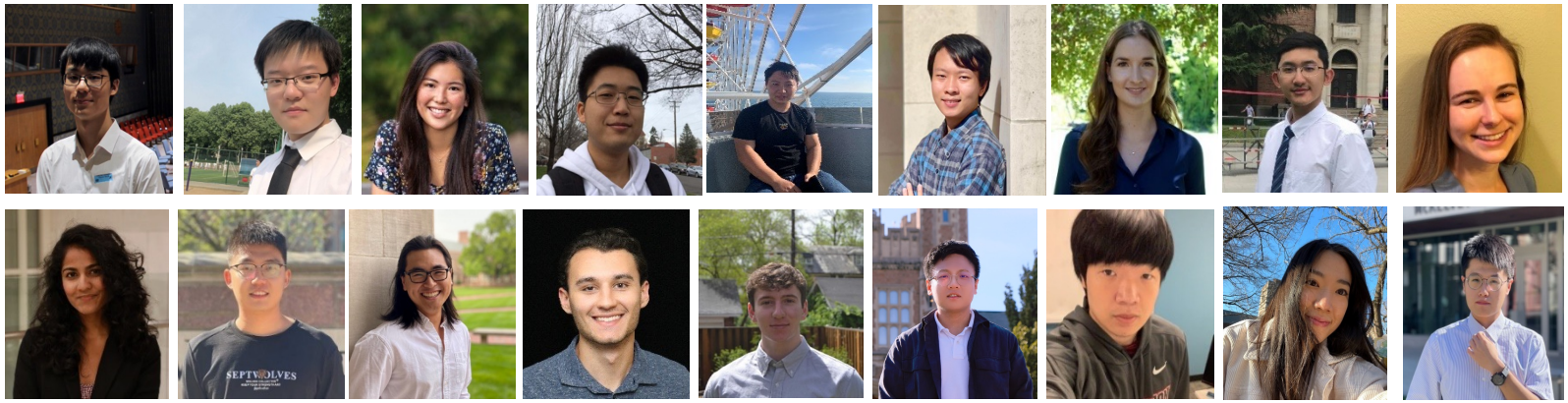
- Results
  - 39 attacked models: 37 detected, 2 false negatives
  - 8 benign models: 6 true negatives, 2 false positives





# Many Thanks !

CSPL



NSF

U.S. DEPARTMENT OF HOMELAND SECURITY

SCIENCE TO SHAPE THE FUTURE OF THE ARMY · ARMY RESEARCH OFFICE · 1951

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

WASHINGTON UNIVERSITY IN ST. LOUIS · Per Veritatem Vis · 1853