Note: Although some of the omitted DEFCON China live presentation slides have been restored in this upload instance, some of the content, specifically regarding military systems and genetic research, has been left out.

Find me on twitter @F1F1cin and message me for the original sources or to comment on the omitted sections of the talk

A Bit About Me

(I‘m going to pretend you care)

F1F1cin

Student at Columbia University in New York

Independent Researcher

Mostly focus on malware

Probably younger than you think

I want to hack a human one day (judge all you want)

Machine Learning

Machine Learning as a Tool

Machine Learning as a Tool for Societal Exploitation

A Summary on the Current and Future State of Affairs

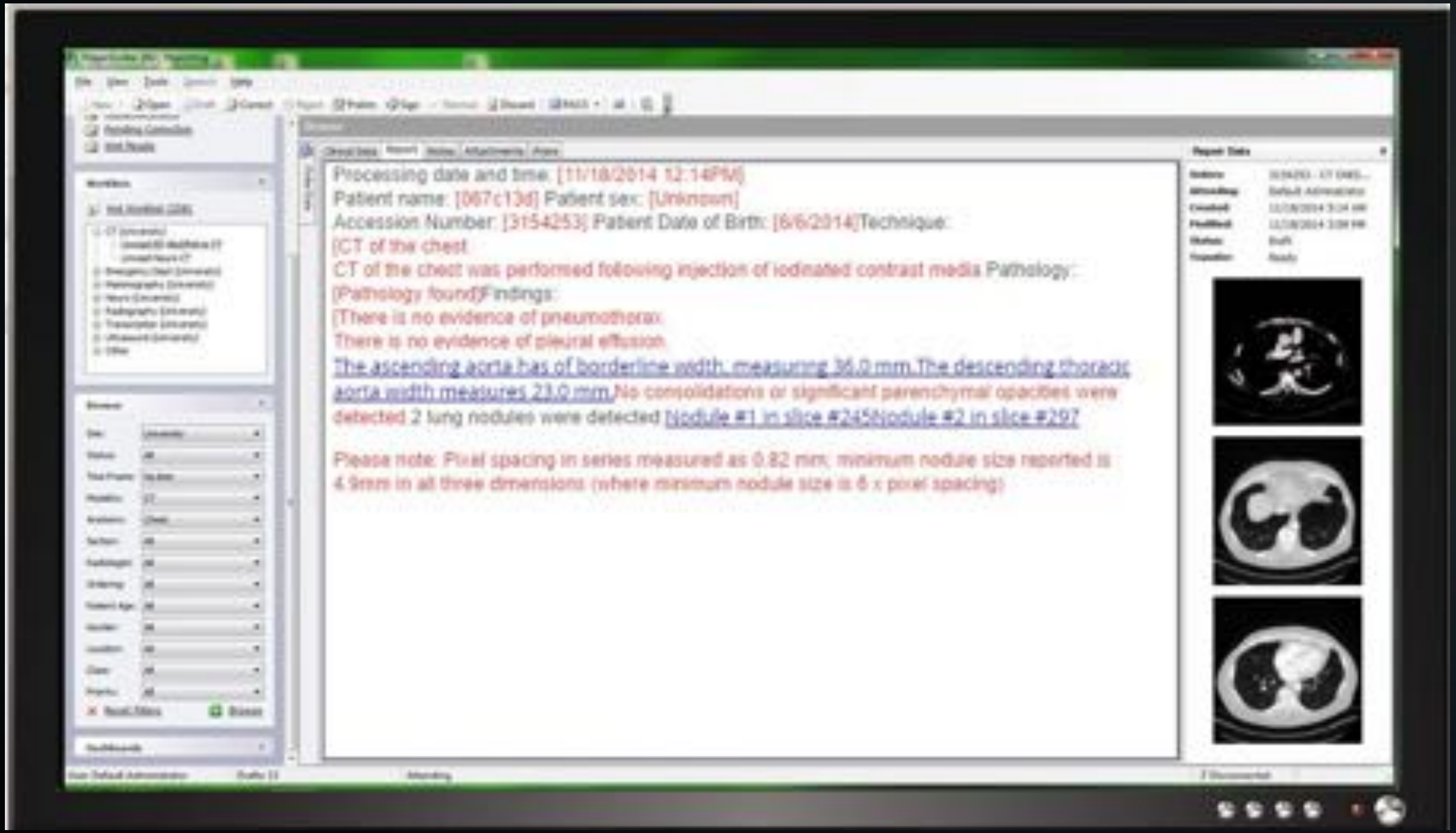
Current State

The Common and the Uncommon

Standard Uses

(generally beneficial, sometimes concerning)

The „Human“ Side



Processing date and time: [11/18/2014 12:14PM]
Patient name: [067c13d] Patient sex: [Unknown]
Accession Number: [3154253] Patient Date of Birth: [6/6/2014]Technique:
[CT of the chest
CT of the chest was performed following injection of iodinated contrast media Pathology:
[Pathology found]Findings:
[There is no evidence of pneumothorax
There is no evidence of pleural effusion
The ascending aorta has of borderline width, measuring 36.0 mm.The descending thoracic aorta width measures 23.0 mm.No consolidations or significant parenchymal opacities were detected 2 lung nodules were detected Nodule #1 in slice #245Nodule #2 in slice #297

Please note: Pixel spacing in series measured as 0.82 mm; minimum nodule size reported is 4.9mm in all three dimensions (where minimum nodule size is 6 x pixel spacing)

Financial Trading

Sports Injuries – [courtesy of Quantum Black]

Medical Imaging

Image from RADLogics

The „Technical“ Side

Data Security

Antivirus Software

Endpoint Detection Systems

The „Technical“ Side

Data Security

Antivirus Software

Endpoint Detection Systems

Uncommon Uses

(usually concerning, generally cool)

REALLY

Crazy Dystopian S**t

Ambient Sound Mapping

- Determine precise location and orientation through microphone-embedded devices [without consent]

Individual Profiling

- Recreating the human based on digital fingerprints

Ambient Sound Mapping

(Initially) Geospatial ambient sound mapping through the NGI

Adopted by a start-up who wants to make people easier to find

Live training sets (live modification)

Crazy Dystopian S**t

Ambient Sound Mapping

- Determine precise location and orientation through microphone-embedded devices [without consent]

Individual Profiling

- Recreating the human based on digital fingerprints
- Actually more common than I give it credit for

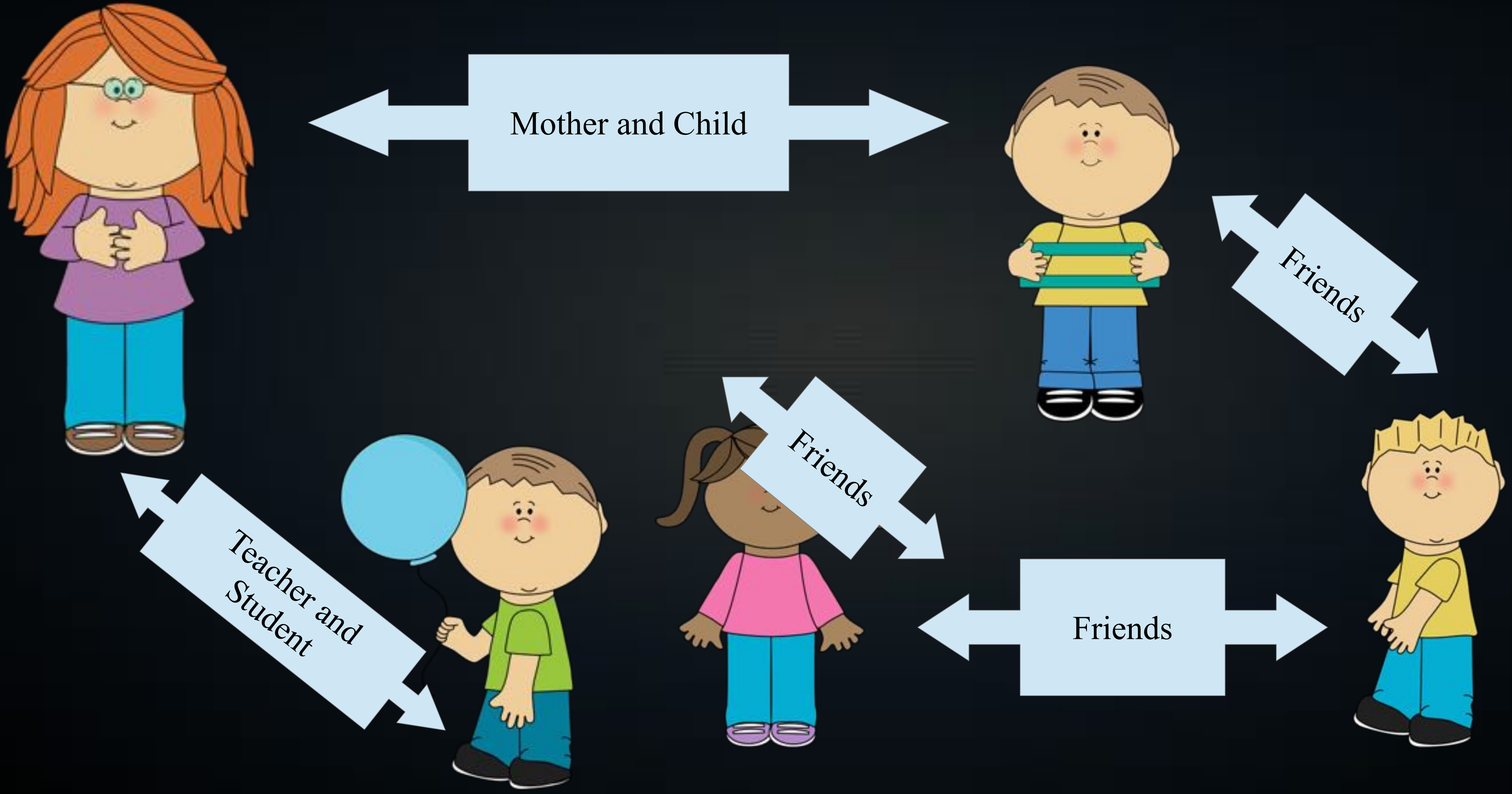
Individual Profiling

Companies take in a lot of information from their cutomers

Facebook, Google, even some hardware manufacturers: take information of their users and construct a type of fingerprint

Even if there is no photo or name, your fingerprint remains the same. A „change“ in identity is not enough.

Individual Profiling



Mother and Child

Friends

Friends

Teacher and Student

Friends

The Future of Attack

FIRST THING TO REMEMBER

AI is NOT Attackproof

(I‘m sure you know this)

AI is NOT Attackproof

„Attack“ isn‘t limited to using AI as a weapon

AI is NOT Attackproof

„Attack“ isn‘t limited to using AI as a weapon

„Attack“ can mean attacks targetted towards AI systems

AI as a Weapon

Current Experiments / Research / whatever you want to call it

„Whatever you want to call it“

Wargames – [courtesy of Endgame], and the case of the 2014 DEFCON black badge-holding machine (Cyber Grand Challenge)

Intelligent Malware

Adapting to a changing environment

Attacks on AI Systems

This is not what I Typically Do

BUT

This is not what I Typically Do

BUT

Accidentally joining an AI-based IDS research group drags you into things

This is not what I Typically Do

BUT

Accidentally joining an AI-based IDS research group drags you into things

Saying you‘re interested in malware makes people think you write it for fun (and no profit)

This is not what I Typically Do

BUT

Accidentally joining an AI-based IDS research group drags you into things

Saying you‘re interested in malware makes people think you write it for fun (and no profit)

So you‘re put in the attack/testing team

This is not what I Typically Do

BUT

Accidentally joining an AI-based IDS research group drags you into things

Saying you‘re interested in malware makes people think you write it for fun (and no profit)

So you‘re put in the attack/testing team, and then you realize you actually like it

What Can We Do?

The research scenario and its limitations

What Can We Do?

The research scenario and its limitations

- Bureocratic process
- Funding
- Concerns about adversarial research

What Can We Do?

The research scenario and its limitations

Let‘s remember things that happened throughout the weekend.

What Can We Do?

The research scenario and its limitations

Let‘s remember things that happened throughout the weekend.

–Training set reconstruction

.Attacks that exploit probablility

–Adversarial inputs

.EVIL DOTS

–Live memory patches

.A white box attack

What Can We Do?

The research scenario and its limitations

Let‘s remember things that happened throughout the weekend. (and things coming up)

What else can be treated in a similar manner?

Attacking the Human

(one of my goals, but kind of far-fetched at the moment)

Attacking the Human

(one of my goals, but kind of far-fetched at the moment)

Are Humans Simply Complex Algorithms?

Genetic reading for the sake of lactose intolerance

How much of this can be reconstructed for other purposes?

Using the same principle of medical imaging, is it possible to modify a person and their actions or thought processes?

The Future of Defense

Autonomous Military Systems

„Army of None“

Machine Learning-based models of missle defense mechanisms

Problems with testing

Tricking AI in Practice

(and why this is important for defense mechanisms)

Purpose

Tricking AI often depends on knowing what information the algorithm values

The algorithm is simply a reflection of its creators

If you understand the algorithm, you understand what your attacker values

Malware analysis, military countermeasures (and the allocation of reseources)

Military Countermeasures

Note: Slide omitted in original DEFCON China presentation

Recreating the training set to learn about the enemy

Important variables (strong weight, multiple instances, etc) can be used to identify points of interest

A more focused adversary can use this information to create a working „fooling“ technique without having to deal with many technical modifications

The Overlaps

You might notice overlaps between attack and defense

Like any other tool, AI can be used on both ends of the spectrum, sometimes without much modification

Defense for the Common Man

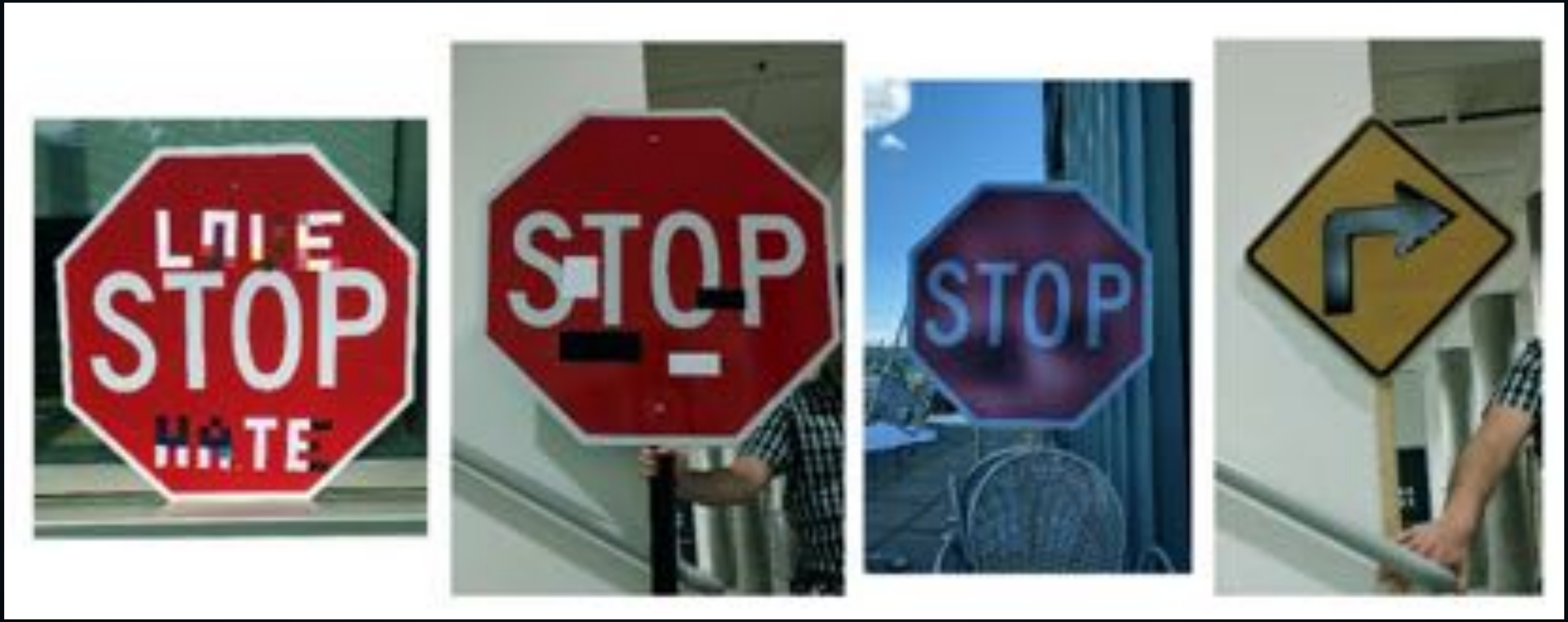
…

(Attack against the algorithm)

A Sample of Defense :

Avoiding Identification and the creation of a false fingerprint

We Have Seen This Before



We Have Seen This Before

https://vimeo.com/208642358

Demo time ?