RSAConference2019
Asia Pacific & Japan
Singapore | 16–18 July | Marina Bay Sands

SESSION ID: HPS-W01

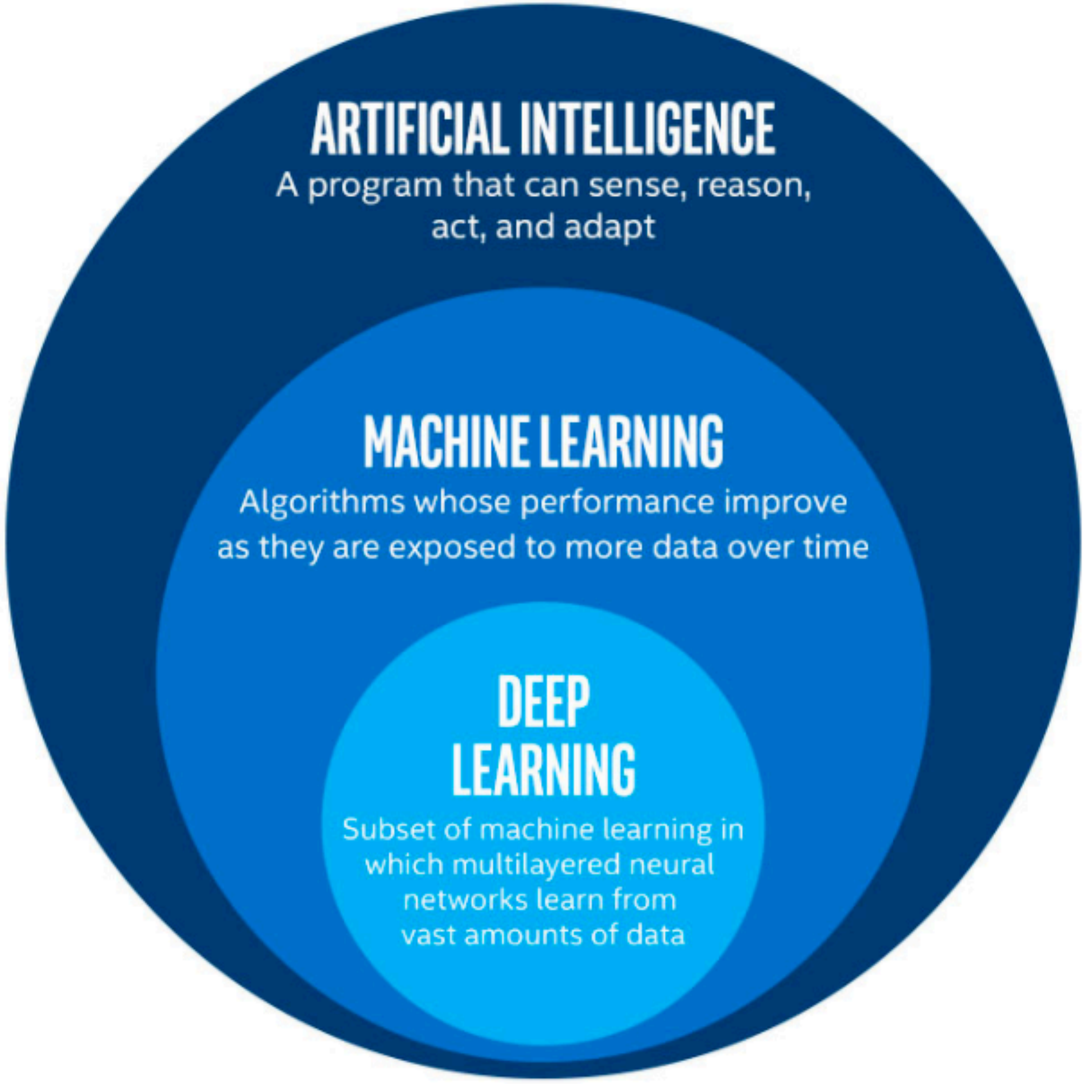
Artificial Intelligence in Cybersecurity: the Good, the Bad and the Ugly

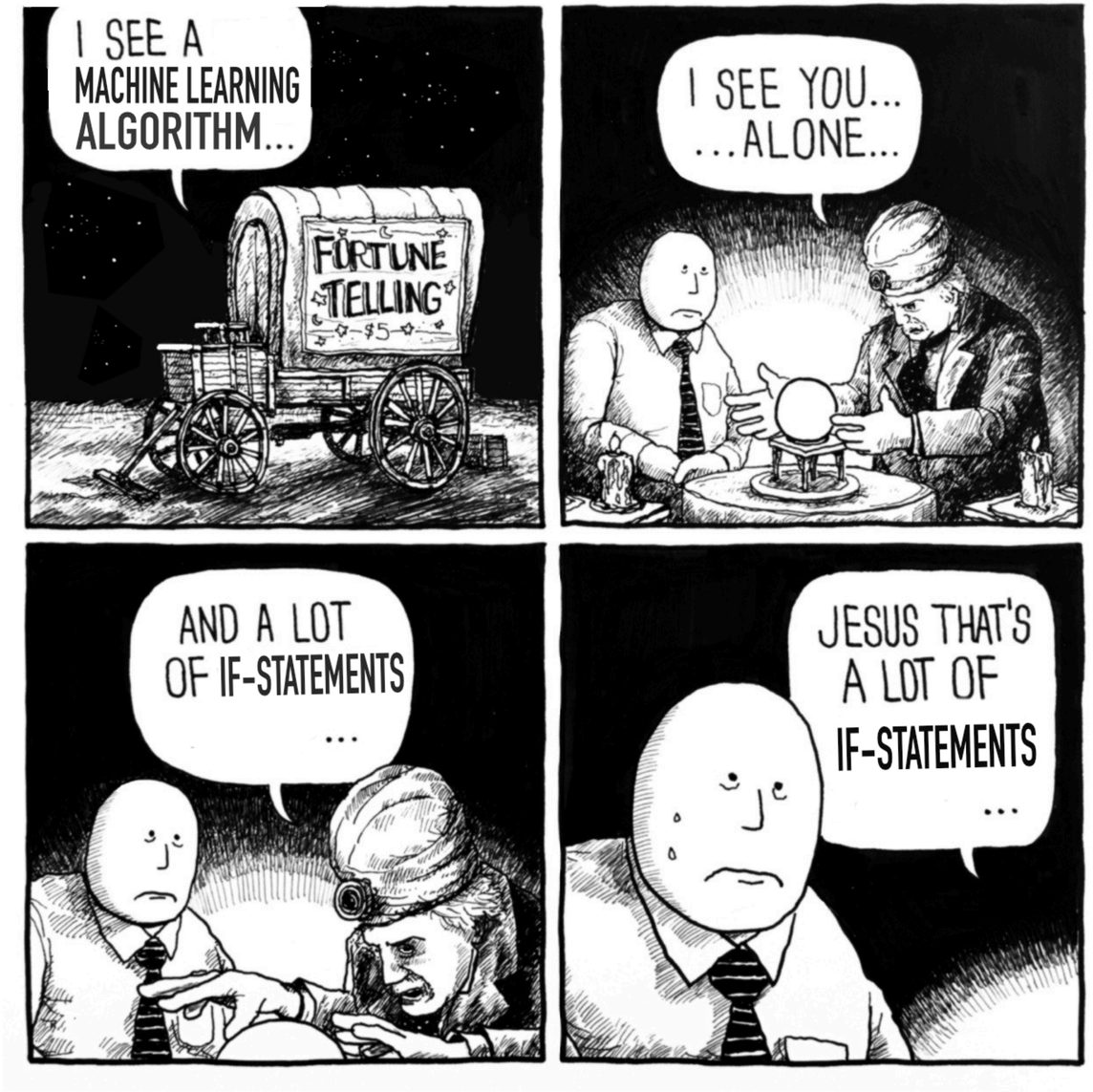
Aleksandr Lazarenko,
Head of R&D Department
Group-IB
@lazarenkoale

WHOAMI

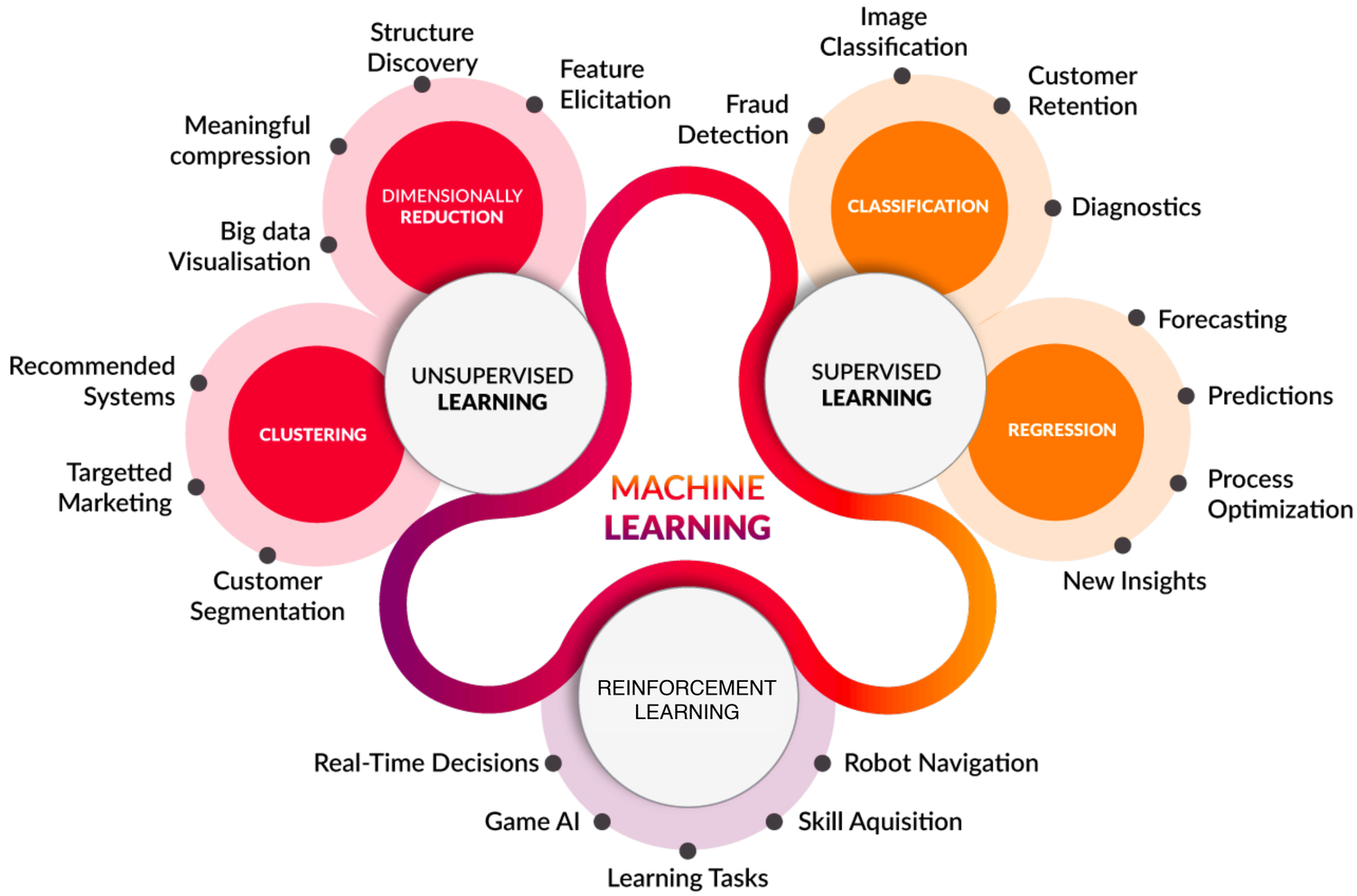
- Aleksandr Lazarenko

- Head of R&D Deptartment
- Former Cyberdetective
- BSc, Software Engineering
- Author of scientific papers on privacy and anonymity, security of blockchain projects
- Inventor of patented technologies


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data


I SEE A MACHINE LEARNING ALGORITHM...
FORTUNE TELLING
$5
I SEE YOU...
...ALONE...
AND A LOT OF IF-STATEMENTS
...
JESUS THAT'S A LOT OF IF-STATEMENTS
...

Types of Machine Learning



The rise of AI/ML in cybersecurity

Forbes CommunityVoice Connecting expert communities to the Forbes audience. What is This?

2,519 views | Nov 9, 2018, 07:30am

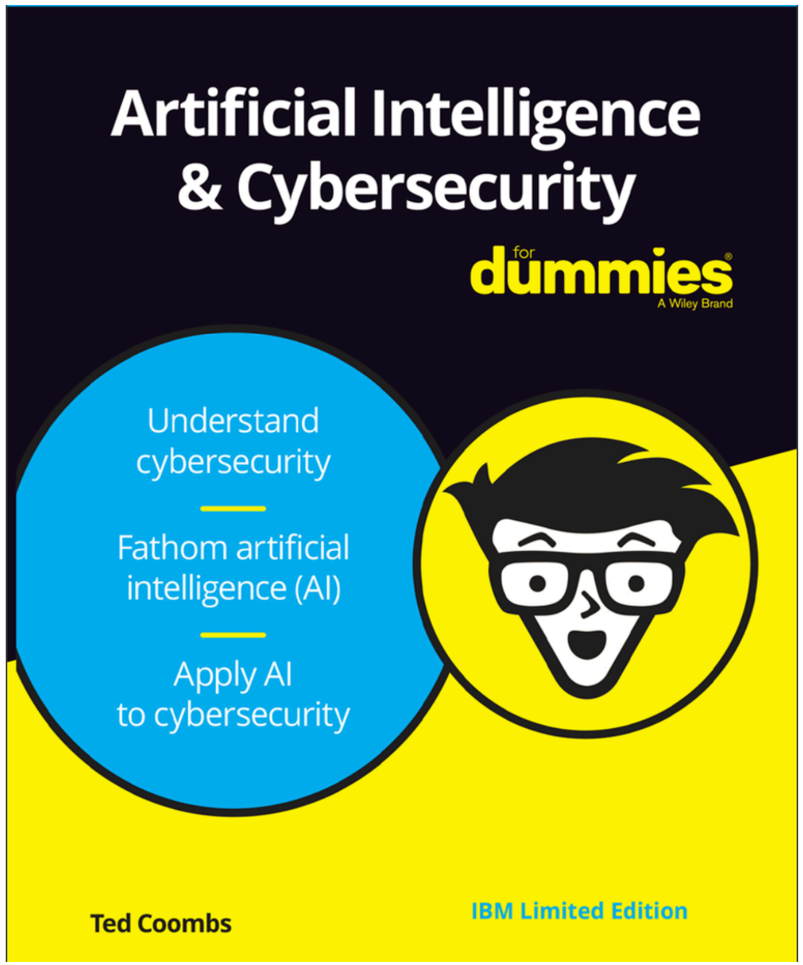
The Rise Of The Intelligent Machine In Cybersecurity

Forbes Technology Council

Bret Piatt Forbes Councils
Forbes Technology Council CommunityVoice

RELATED EXPERTISE: TECHNOLOGY & DIGITAL

Artificial Intelligence Is a Threat to Cybersecurity. It's Also a Solution.



34%

Report their organisation has experienced a damaging cyberattack in the last 12 months

72%

Agree that as long as their protection keeps them safe from cybercriminals, they do not care if it uses AI/ML

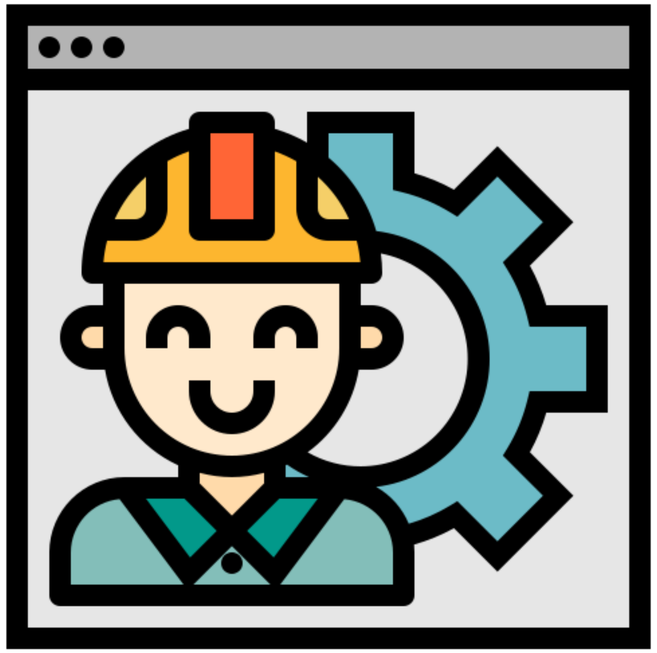
73%

Plan to use more AI/ML tools in 2019

84%

Believe cybercriminals are using AI/ML to attack organisations

Who is going to win the battle?



Cybersecurity professionals

VS

Hackers

RSAConference2019
Asia Pacific & Japan

How to use AI/ML in cyberattacks?

The Bad

DARPA-sponsored Cyber Grand Challenge

Mayhem got the ticket to DEF CON CTF!

One of the proudest moments of my career, years ago, was the chance to be part of a team that earned entry to DEF CON Capture the Flag.

Today, full automation will enter the most competitive hacking contest on Earth as a machine enters this competition for the first time," Walker said.

"I don't expect Mayhem to finish well. This competition is played by masters and this is their home turf. Any finish for the machine save last place would be shocking.

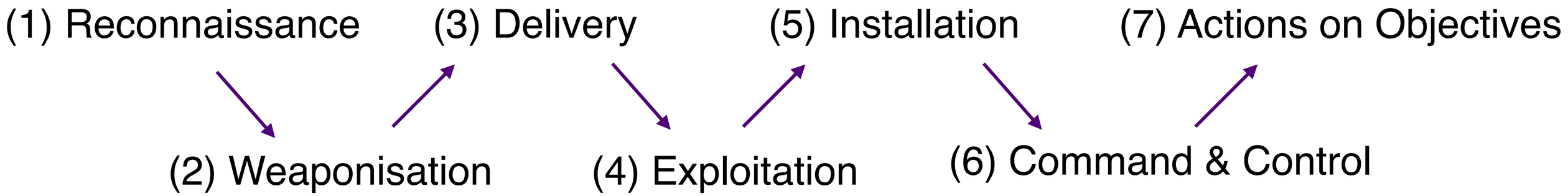
(Mayhem finished last….)

Mayhem @ DEF CON CTF

Team	Final Score
PPP	113555
b1oOp	98891
DEFK0R	97468
HITC0N	93539
KaisHack GoN	91331
LC↯BC	84412
Eat Sleep Pwn Repeat	80859
binja	80812
pasten	78518
Shellphish	78044
9447	77722
Dragon Sector	75320
!SpamAndHex	73993
侍	73368
Mayhem	72047

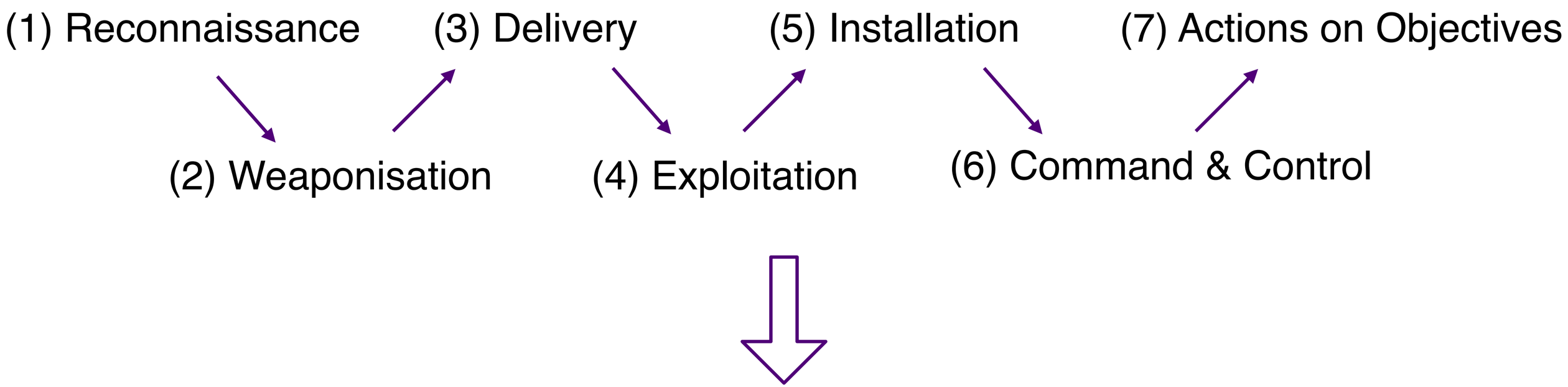
Congratulations to our top three teams PPP, b1oOp, and DEFK0R. We would also like to congratulate all competing teams for spectacular performances all around. This year's game was a drastic departure from previous DEF C0N CTF games, and we appreciate the sacrifices you made to compete in it. Finally, we would in particular like to congratulate Mayhem, from For All Secure, for their spectacular performance as the first autonomous computer system to play DEF C0N CTF. While Mayhem did finish in last place, many times throughout the game it was able to pull ahead of human teams.

Let’s look at the Cyber Kill Chain Model



https://www.sans.org/security-awareness-training/blog/applying-security-awareness-cyber-kill-chain

Let's look at the Cyber Kill Chain Model



All these tasks could be enforced with AI or ML tools!

https://www.sans.org/security-awareness-training/blog/applying-security-awareness-cyber-kill-chain

(1) Reconnaissance

People

- Information classification via NLP
- Characterising potential victims
- Finding victims solvency
- Detect social media via image recognition
- Advanced SNA to find influencers in communities
- Tracking people through wearables
- Searching for sensitive information in SN

IT Infrastructure

- Analysing SDN structure (KYE attack)
- Probing traffic correlation
- ML for directory name generation
- Network Big Data analysis

Free Facial Recognition Tool Can Track People Across Social Media Sites

August 09, 2018 Swati Khandelwal

(2) Weaponisation && (3) Delivery

- Fake news content generation
- Content analysis
- Target analysis
- Tailored phishing campaigns
- Content distribution
- Impersonation in Spam
- Impersonation in Phishing

(4) Exploitation && (5) Installation

As a Tool

- Vulnerability discovering
- Attack fitting
- Human-like denial of service
- CAPTCHA bypass
- Password brute force

For malware

- Samples generation
- Mimics
- Features adjustment
- Self-destruction
- Hivenets

PassGAN: A Deep Learning Approach for Password Guessing*

Briland Hitaj
Stevens Institute of Technology
bhitaj@stevens.edu

Paolo Gasti
New York Institute of Technology
pgasti@nyit.edu

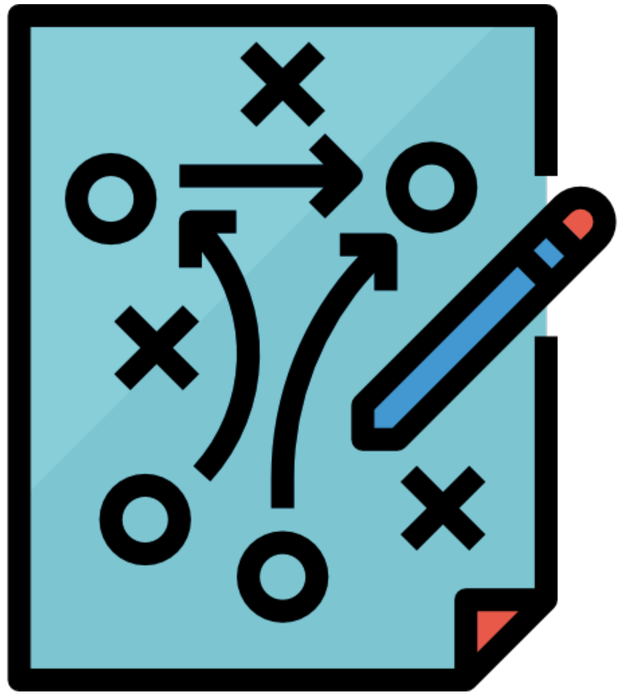
Giuseppe Ateniese
Stevens Institute of Technology
gatenies@stevens.edu

Fernando Perez-Cruz
Swiss Data Science Center, (ETH Zurich and EPFL)
fernando.perezcruz@sdsc.ethz.ch

I'm not a human: Breaking the Google reCAPTCHA

Suphannee Sivakorn, Jason Polakis, and Angelos D. Keromytis
[suphannee, polakis, angelos]@cs.columbia.edu
Columbia University, New York NY, USA

(6) C&C && (7) Actions on Objectives



- Tactics/strategy optimisation
- Target selection
- Target assessment
- Attack planning
- Complex decision making

The first cyberattack using AI

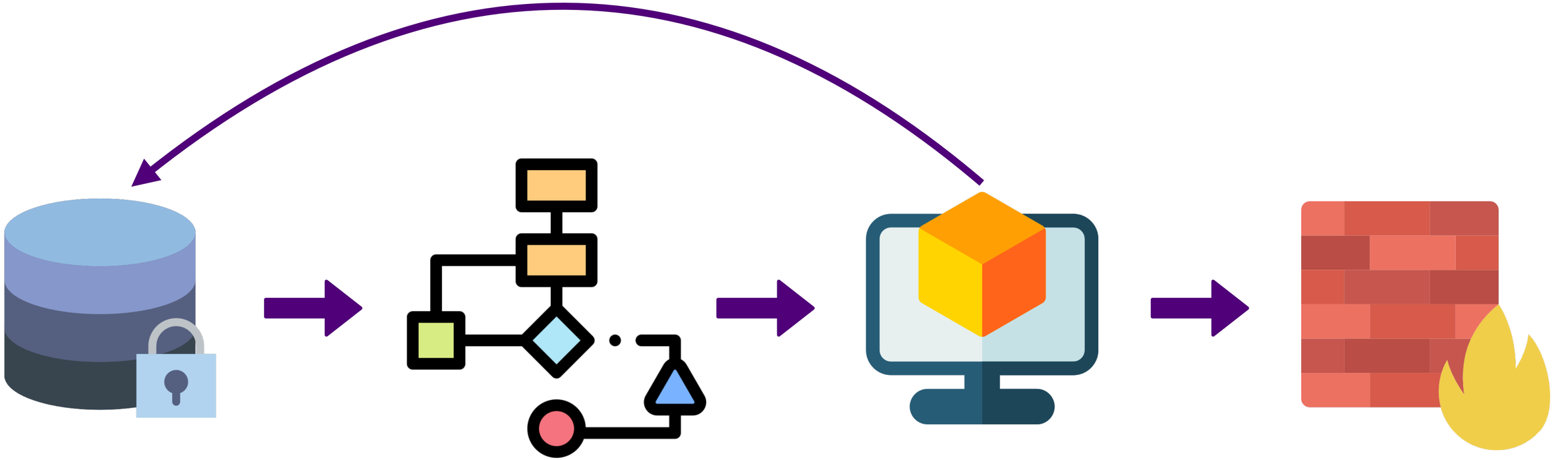
- Rudimentary machine learning to observe and learn patterns of normal user behavior inside a network
- The software began to mimic behavior, blending into the background and becoming harder for security tools to spot

FORTINET®

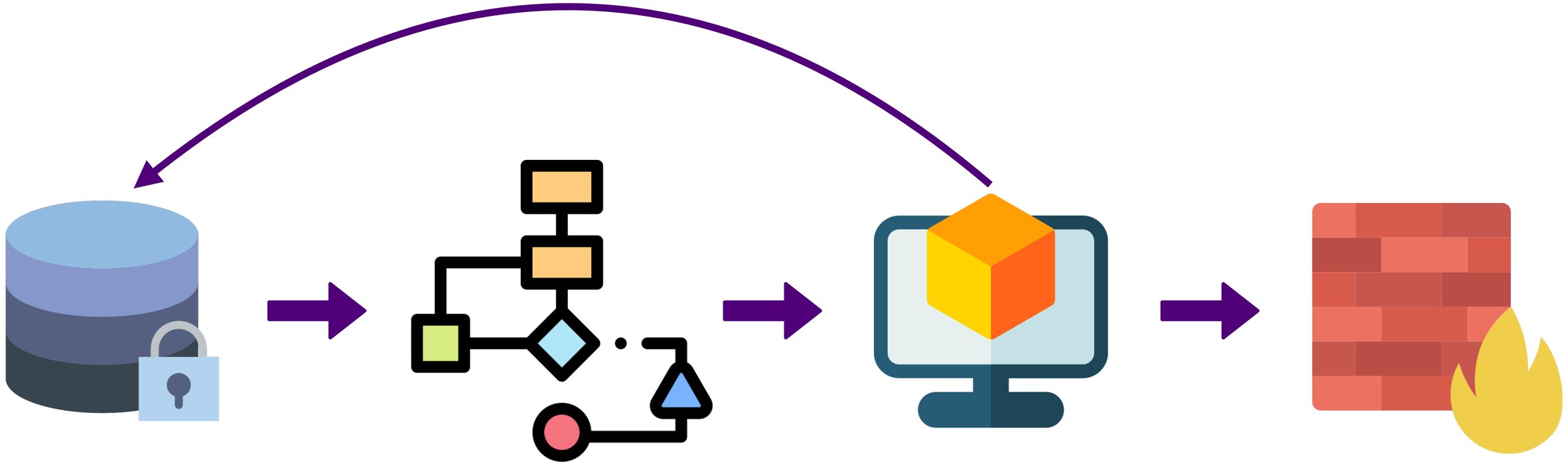
“We do imagine that there will be a time when attackers use machine learning and artificial intelligence as part of the attack. We have seen early signs of that,”

Chief Executive Nicole Eagan (Darktrace Inc.)

Hacking the AI-based system?



Hacking the AI-based system?



Datasets

- Hacking into the datasets used for training and altering the data in such a manner that it will acquire poisoned properties

Training algorithm

- Altering the code of an algorithm

Trained model

- Adversary can “try” to mine properties of model with different queries
- Poisoning via biases

Security system

- Tailored requests to defeat ML in security system

Attacks taxonomy

- INFLUENCE
  – Causative attacks influence learning with control over training data.
  – Exploratory attacks exploit misclassifications but do not affect training.
- SECURITY VIOLATION
  – Integrity attacks compromise assets via false negatives.
  – Availability attacks cause denial of service, usually via false positives.
- SPECIFICITY
  – Targeted attacks focus on a particular instance.
  – Indiscriminate attacks encompass a wide class of instances.

RSAConference2019
Asia Pacific & Japan

How can we use AI/ML to protect our business from cyberattacks?

The Good

Already implemented…

- To identify anomalies;
- To identify suspicious or unusual behaviour;
- Detect and correct known vulnerabilities;
- Detect and correct suspicious behaviour;
- Detect and correct zero-day attacks
- And almost any other cybersecurity product powered by AI or ML…

Types of solutions which use AI

- Anti-fraud & Identity Management
- Mobile Security
- Predictive Threat Intelligence
- Behavioural Analysis & Anomaly Detection
- Automated Security
- Cyber-Risk Management
- App Security
- IoT Security
- Deception Security
- Malware classification
- Spam identification
- DNS analytics
- Analyst automation

When things come bad?

- Not enough or no quality labelled data
- “Dirty” data
- Bad understanding of the data

CYBERSECURITY'S NEXT STEP MARKET MAP:
80+ COMPANIES SECURING THE FUTURE WITH ARTIFICIAL INTELLIGENCE

ANTI FRAUD & IDENTITY MANAGEMENT

AGARI
邦盛科技 Bangsun Technology
Castle
CYBERTONICA
DATAVISOR
feedzai
GreatHorn
GYOMO
id wall
PRECOGNITIVE
Ravelin
pulseiD
rippleshot
Shift Technology
数美 NEXTDATA
smyte.
skymind
simility
sift science
SOCURE
ZYUDLYLABS
CVU
veridu
UNFRAUD
trooly

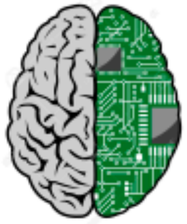
MOBILE SECURITY

appthority
Mi³ Security
Sentegrity
Skycure
ZIMPERIUM.

PREDICTIVE INTELLIGENCE

CYLANCE
deepinstinct
indeni
INNEFU
IntelliSpyre
JASK
LogRhythm
SECLYTICS
SECB
signalsense
ANOMALI
PROTENUS
patternex
SentinelOne

BEHAVIORAL ANALYTICS / ANOMALY DETECTION



Avata INTELLIGENCE
BehavioSec
DARKTRACE
E8 SECURITY
perimeterx
exabeam
SECUREDTOUCH
SPHERICALDEFENCE
RUBICA
RedLock
INTERSET
intensity analytics
FORTSCALE
CyberX
sqrrl
TWOSENSE.AI
StackPath

AUTOMATED SECURITY

DEMISTO
EdgeWave
JAVELIN
LogicHub
TANIUM
ZENEDGE

CYBER-RISK MANAGEMENT

cybersaint
CYENCE
Cytora
Haystax
lexumo
METACERT
wiretap

APP SECURITY

AuthBase
Cryptosense
CYBER 20/20

IOT SECURITY

sparkcognition
Bastille
CUJO

DECEPTION SECURITY

CyberFog
illusive

GROUP IB

CBINSIGHTS

What is the problem?

The Ugly

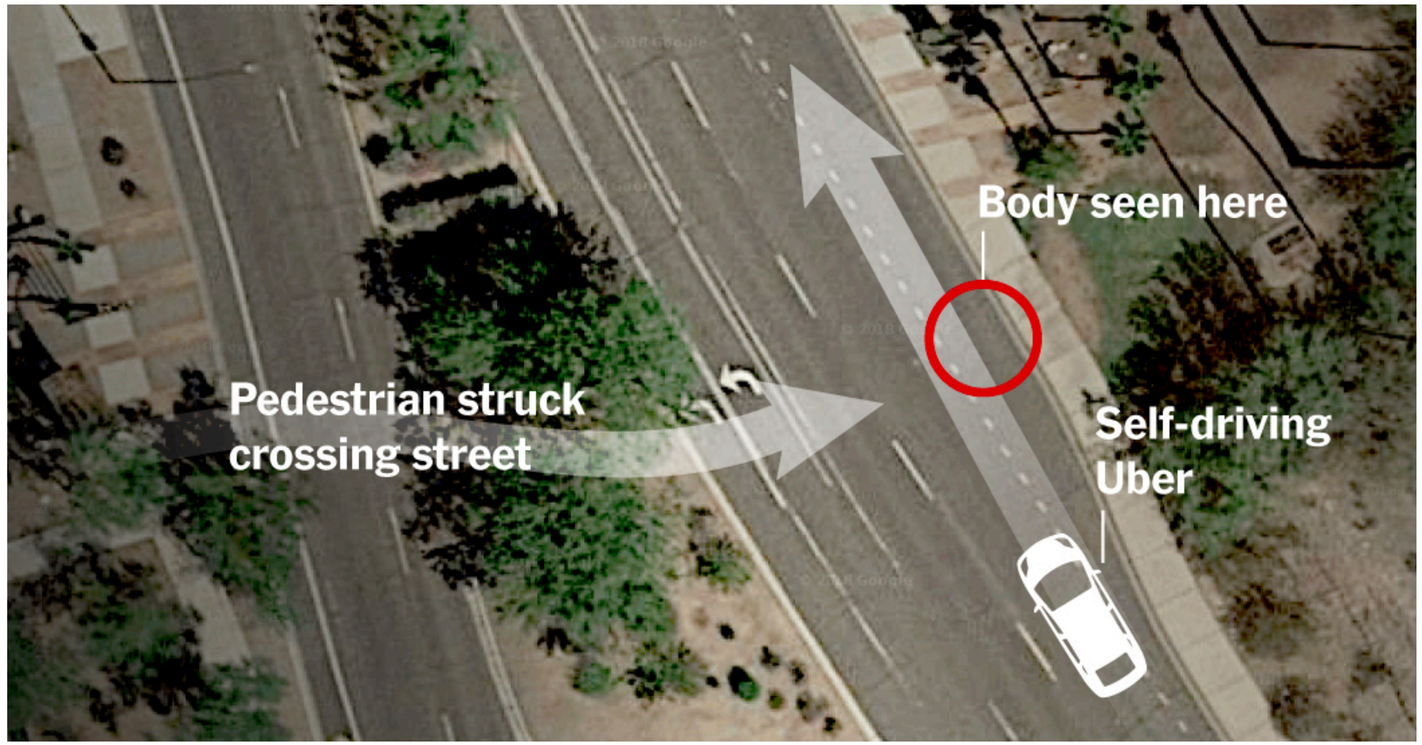
AI fails….

Chinese billionaire's face identified as jaywalker

- Traffic police in major Chinese cities are using AI to address jaywalking.
- The AI system in the southern port city of Ningbo however recently embarrassed itself when it falsely "recognized" a photo of Chinese billionaire Mingzhu Dong on an ad on a passing bus as a jaywalker

Uber self-driving car kills a pedestrian

- An Uber self-driving SUV struck and killed a female pedestrian on March 28 in Tempe, Arizona
- Self-driving software decided not to take any actions after the car's sensors detected the pedestrian.



Amazon AI recruiting tool is gender biased

- Amazon HR reportedly used an AI-enabled recruiting software between 2014 and 2017 to help review resumes and make recommendations
- The software reportedly downgraded resumes that contain the word “women” or implied the applicant was female, for example because they had attended a women’s college

What is the Ugly side of AI/ML?

- Sometimes the algorithms do not learn the right thing but something else
- Testing and debugging is not easy, as we need to deal with a lot of uncertainties.
- Costs of acquisition, operation and maintenance generally related to the highly specialized, scarce and expensive expertise required.
- The impact of regulatory Artificial Intelligence & Machine Learning in Cybersecurity frameworks might be diverse, involving privacy, data protection and other regulations impacting automated decision making.

RSAConference2019
Asia Pacific & Japan

How should we handle the security of AI in security system?

Fundamental Theorem of Security

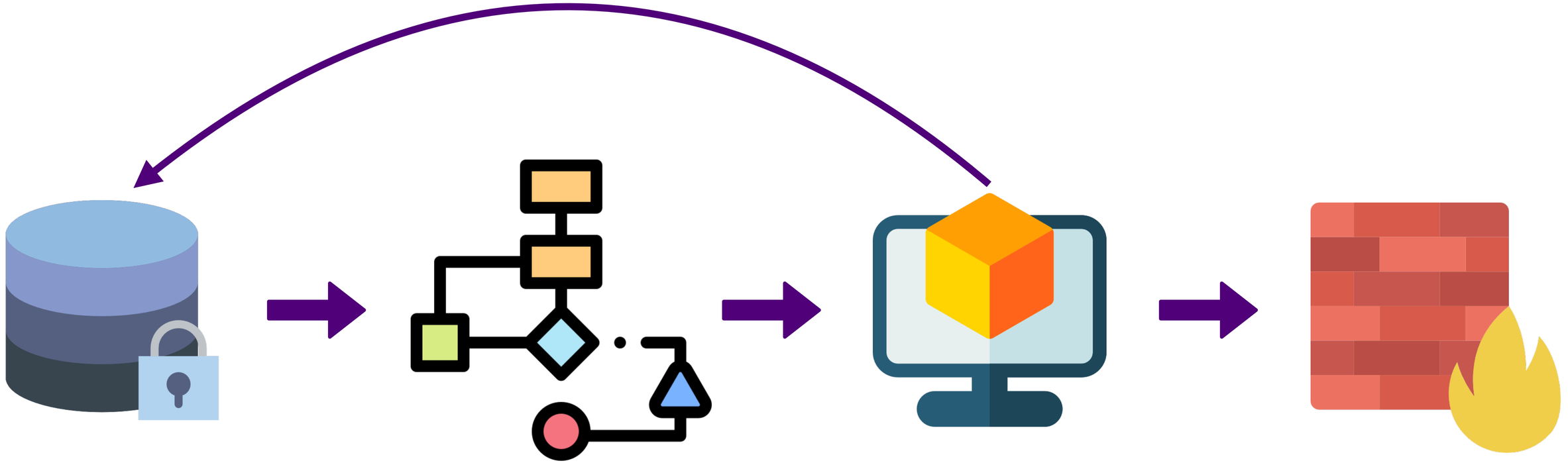
(1)Every security system will eventually fail;
(2)There is no such thing as a 100% secure system.

If your security system has not failed, just wait longer….

Related problems in safety

- Privacy
- Fairness
- Security
- Abuse
- Policy

Protection of key components!



Datasets

- Hacking into the datasets used for training and altering the data in such a manner that it will acquire poisoned properties

Training algorithm

- Altering the code of an algorithm

Trained model

- Adversary can “try” to mine properties of model with different queries
- Poisoning via biases

Security system

- Tailored requests to defeat ML in security system

Machine Learning can do anything you could train a dog to do - but you’re never totally sure what you trained the dog to do.

Apply slide

- Next week you should:
  - Understand the advantages and disadvantages of AI/ML based solutions
  - Build the Threat Model of AI-backed adversary
  - Find out the resources which will keep you up-to-date in the field.

RSAConference2019
Asia Pacific & Japan

Thank you for your attention!

Aleksandr Lazarenko

Head of R&D Department @ Group-IB

lazarenko@group-ib.com