radware

The Artificial Reality of Cyber Defense

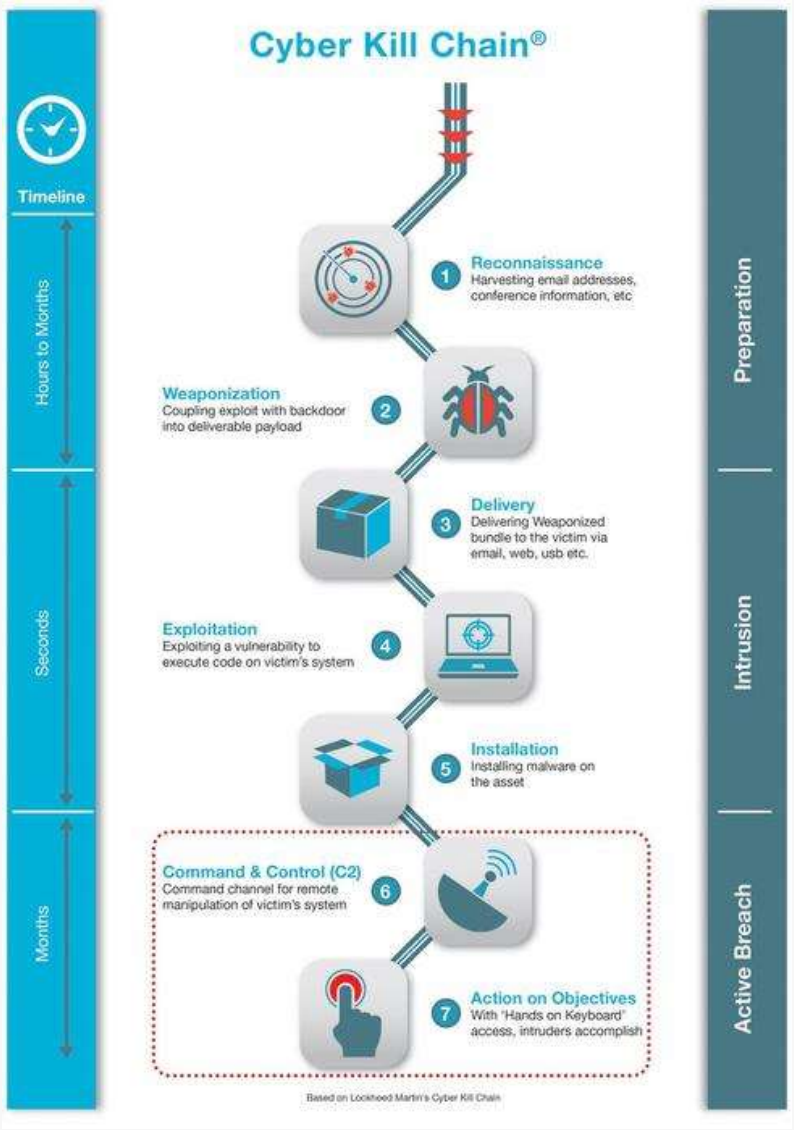
Pascal Geenens

EMEA Security Evangelist

@geenensp

Oct 2018

Cyber Kill Chain® by Lockheed Martin



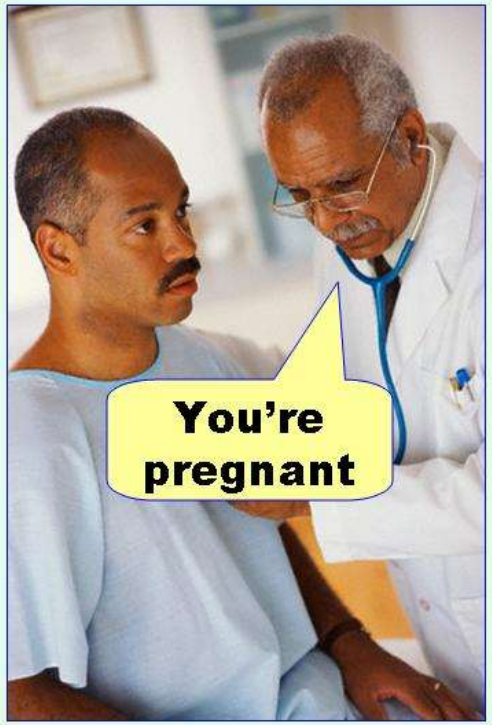
- Targeted attacks
- Plenty of opportunities to detect and block attacks before they cause actual damage

- So why organizations still getting breached and only find out (long) after the fact ; by accident or through ransom ?
- Two reasons mainly:
  - Not enough events / visibility
  - Too many events / false positives

Image Source: Dark Reading - Deconstructing The Cyber Kill Chain - Giora Engel

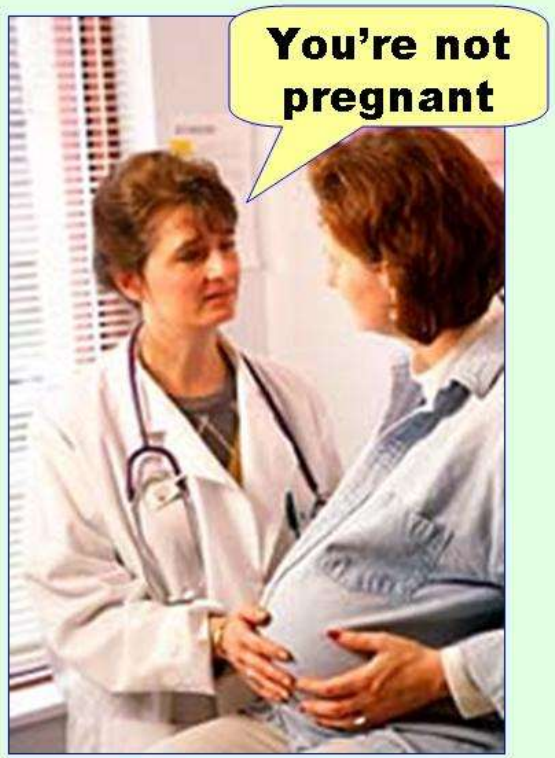
Minimizing False Positives & False Negatives

Type I error
(false positive)



Too many events

Type II error
(false negative)



Not enough events

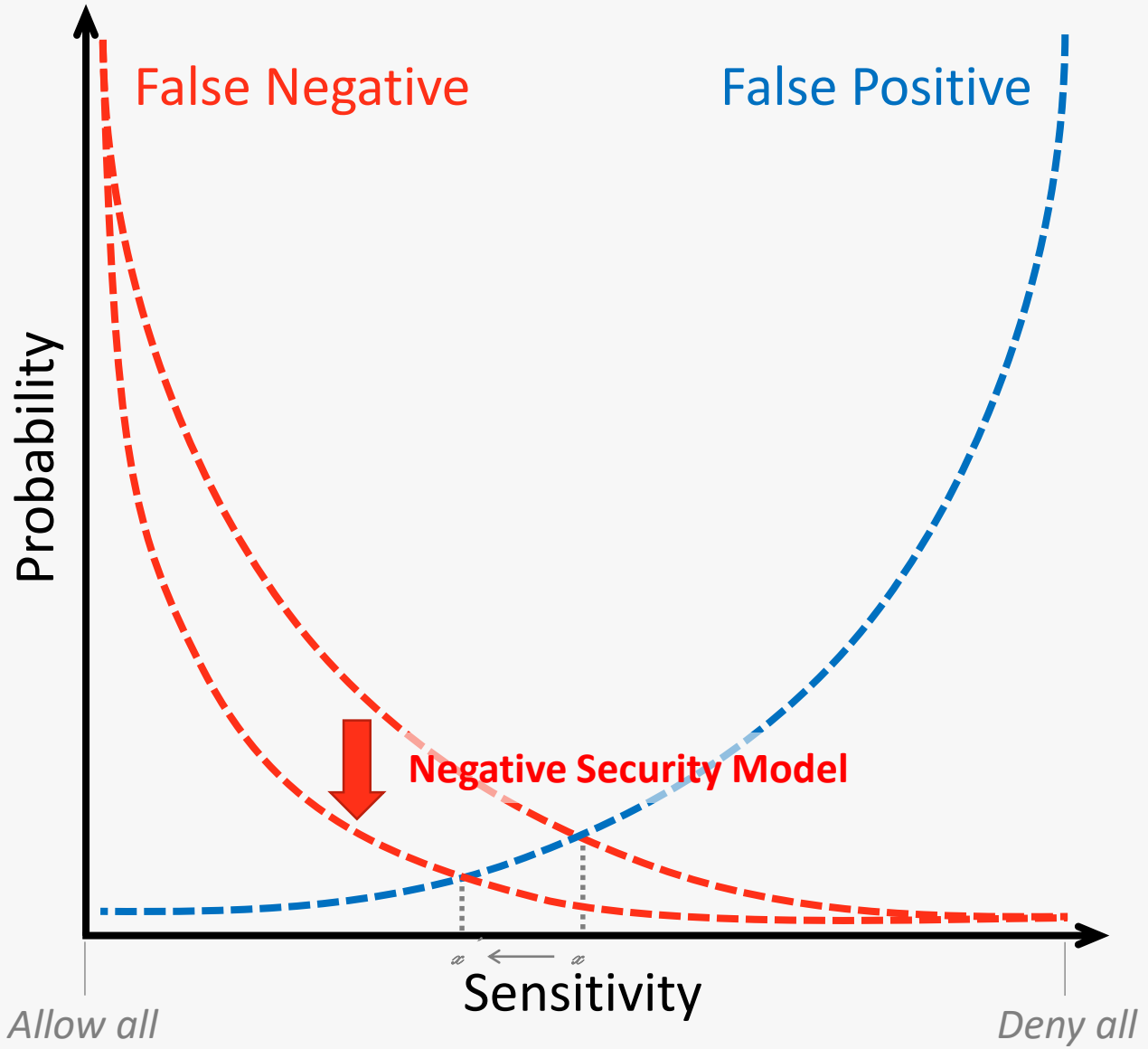
Why minimize

False Negatives?
S3r1ously !?!?

False Positives?
How much incidents can your SOC investigate?
Do you give the right incidents the attention they deserve?

Image Source: Effect Size FAQs by Paul Ellis

Detection Sensitivity in Positive Security Models

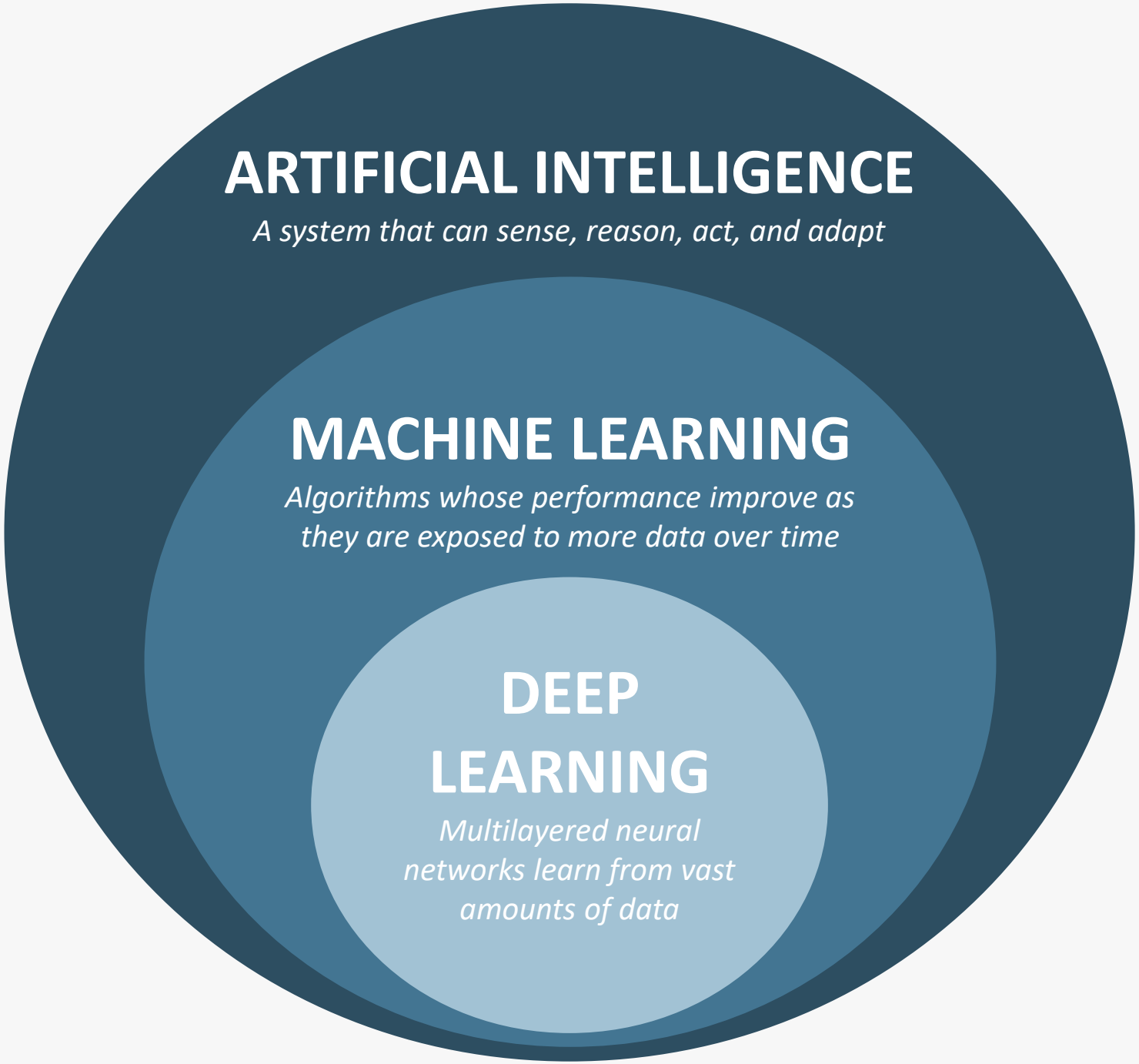


Anomaly Detection – Game On!

- Security threats growing faster than security teams and budgets, huge talent shortage
- Paradox: Proliferation of data from dozens of security products makes it harder to detect and investigate threats

- Rule based event correlation provides reduction from millions to thousands
- A good SOC can investigate maybe a couple of 100 incidents a day
- How to leverage previous work from the SOC to improve the future detection
- Need for automation

Machine Learning


ARTIFICIAL INTELLIGENCE
A system that can sense, reason, act, and adapt
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Multilayered neural networks learn from vast amounts of data

Neural Networks (NN)

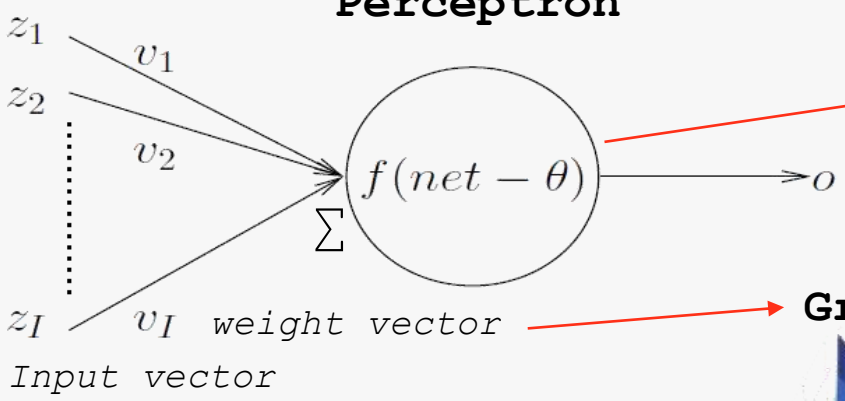
- One class of algorithms in Machine Learning
- Are not mimicking the brain, they are loosely inspired on the brain. They are not even close to what the biological brain does
- Much closer to curve fitting and regression than are to a human brain
- The problem NNs solve is reducing the global error of the output based on known expected output from inputs: optimizing a ‘cost function’

Image Source: https://paulvanderlaken.com/2017/10/16/neural-networks-101/

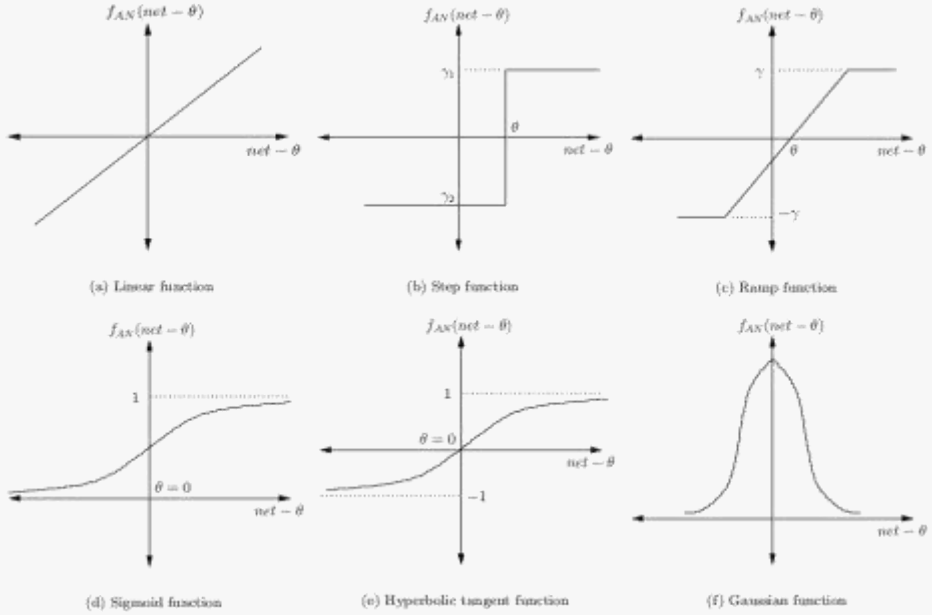
Neural Networks - Concepts

- Objective: minimize some measure of error E for all samples in training data
- Optimization algorithms adjust the weights in the net
- Most common learning algorithm is Gradient Descent

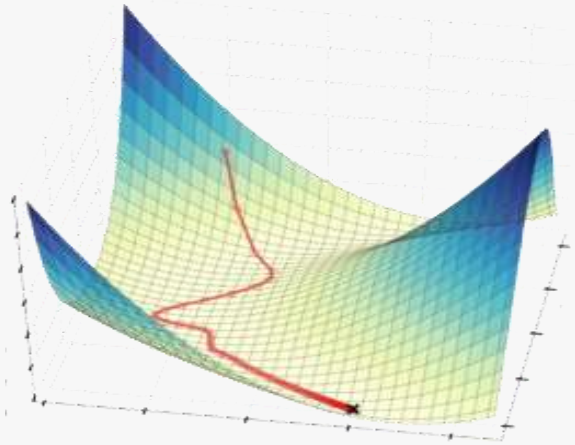
Perceptron



Activation Function



Gradient Descent



Sources: http://www.turingfinance.com/misconceptions-about-neural-networks & http://dsdeepdive.blogspot.com/2016/03/optimizations-of-gradient-descent.html

Deep Learning / Deep Neural Networks (DNN)

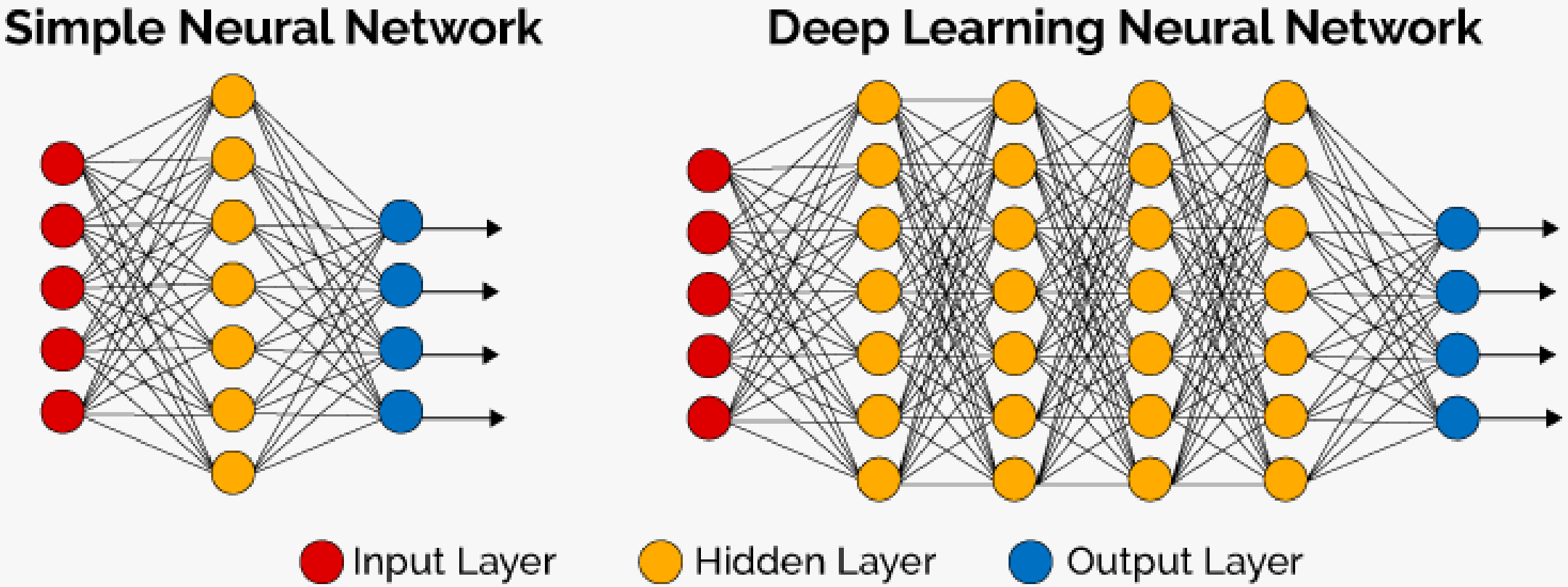
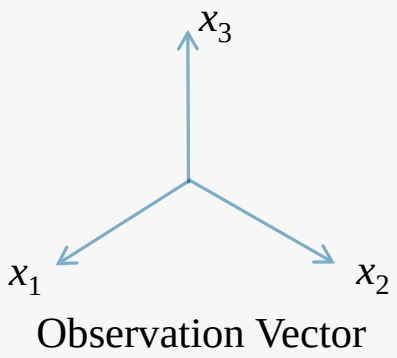
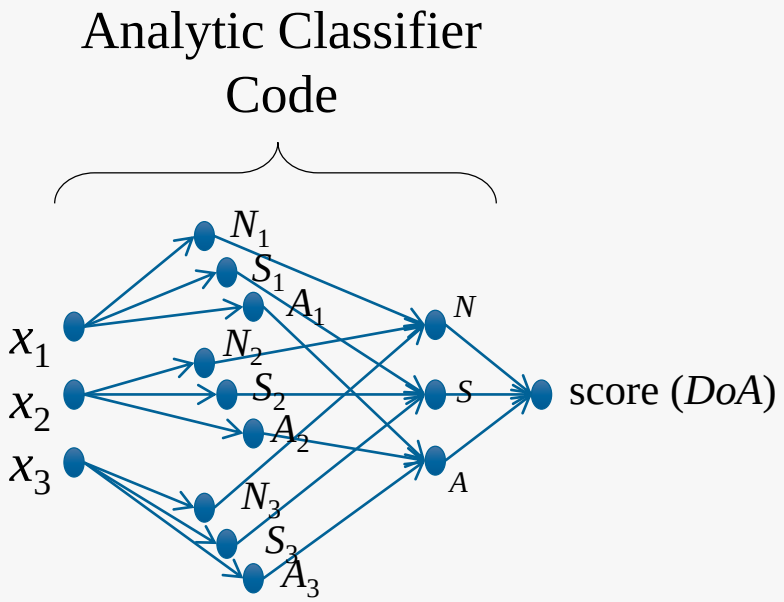
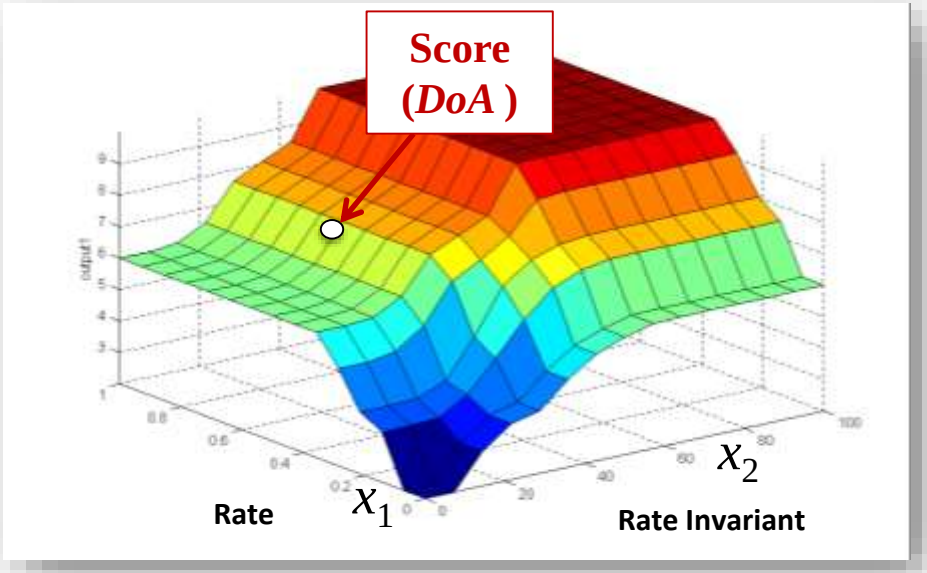


Image Source: http://www.global-engage.com/life-science/deep-learning-in-digital-pathology/

'Traditional' ML - Behavioral-Based Detection Principles

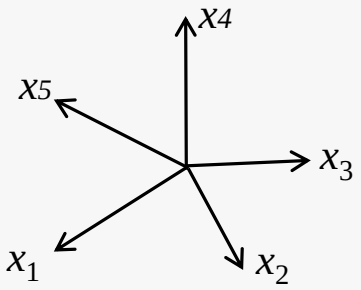




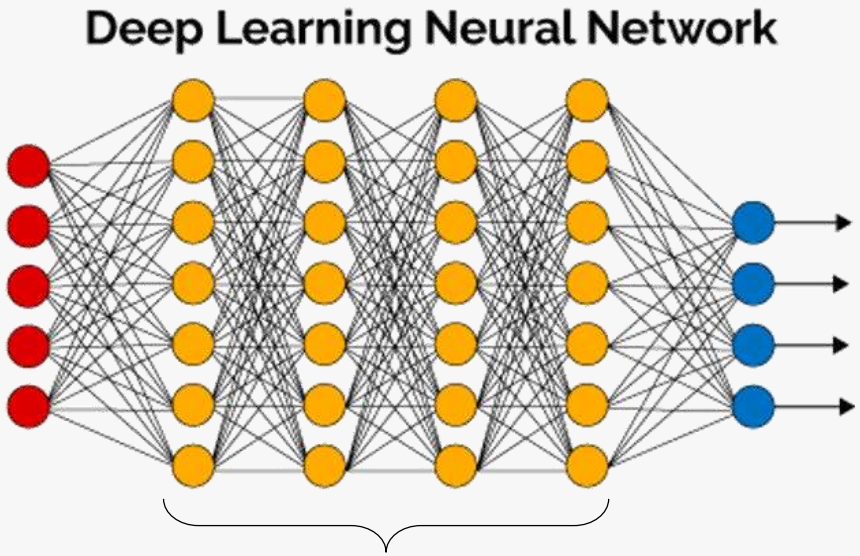


- Complexity of behavioral model is low/med (eg RFC State Machine)
- Code (analytic classifier) can be use to describe the expected behavior
- Data is used for baselining (@ peace-time)
- Limited data sufficient for low false positive rate

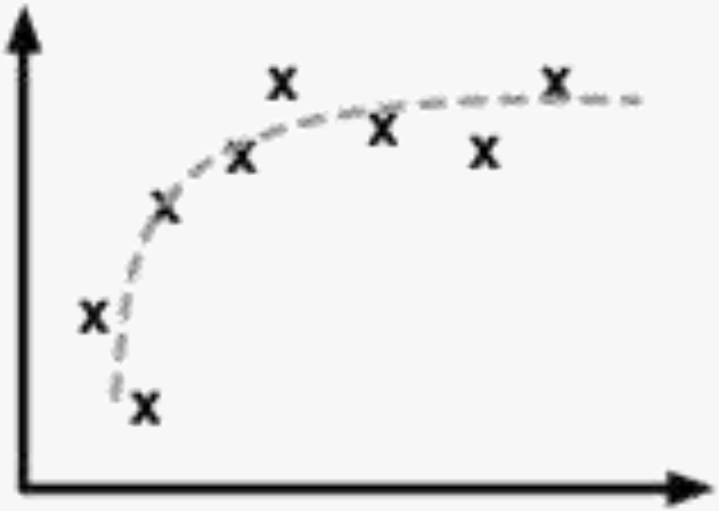
Deep Learning Behavioral Detection Principles



Observation Vector
(e.g Network, Server Logs)



Trained coefficients



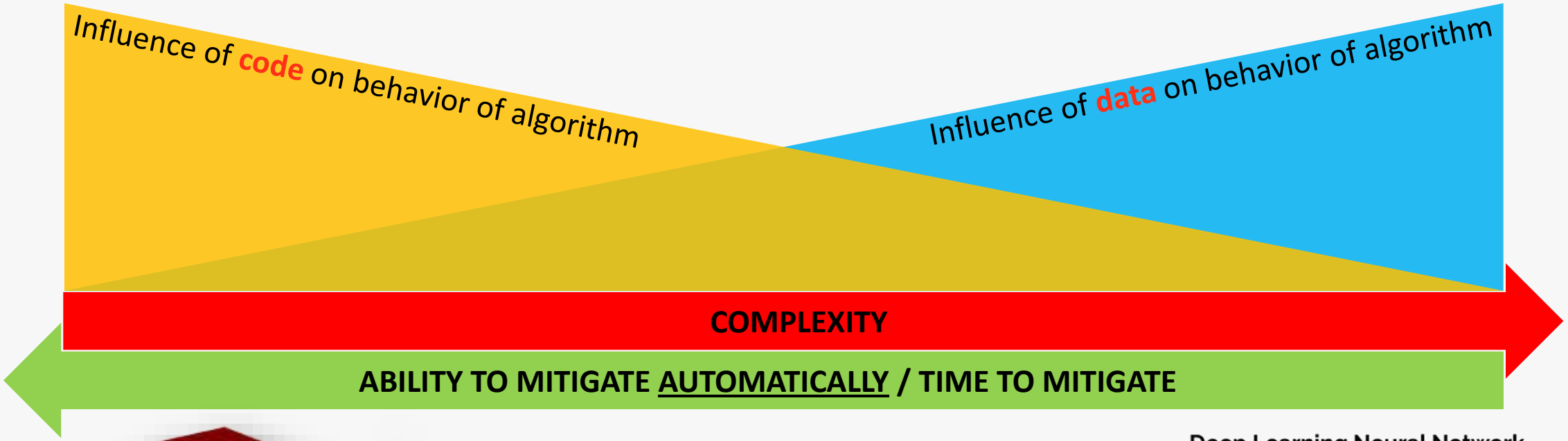
Different Problems Samples

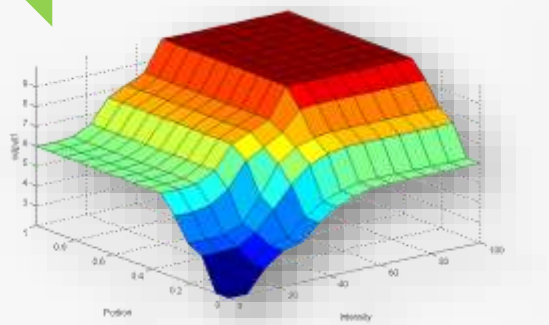
- Complexity of behavioral model is high/very-high
- Can't use code to describe expected behavior
- Data used to describe the expected behavior ("training")
- Lots of 'good' data required

Detection Algorithms & Machine Learning

Deterministic
Transparent
Data provides baselines

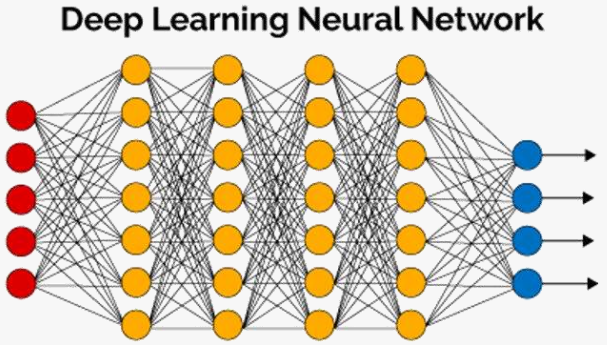
Too complex to code
Generalization
Opaque





Degree of Attack (DoA)

K-means Clustering
Logistic Regression
Bayesian Linear Regression
Support Vector Machine
Principal Component Analysis



Deep Learning Challenges

Challenges of Deep Learning

Training Data

Reproducibility

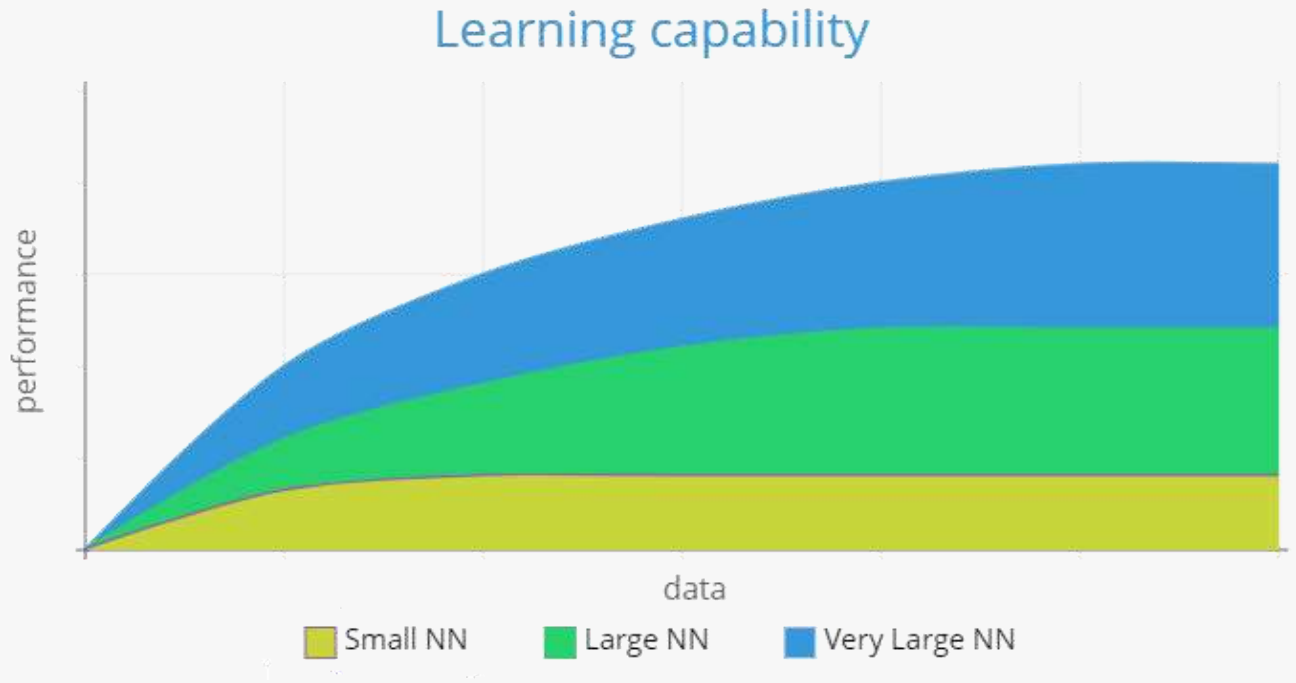
Transparency

Learning in Changing Environments

Learning in Adversarial Contexts

DNNs Need Data! Good Data and Lots of it…

- Larger networks have higher learning capability (memory)
- Performance is only as good as the amount of data put in
- Need extra data to evaluate the network's performance
- Quality of the network will on be as good as the quality of the data put in
- Synthetic data generation can be misleading, correlation between data points



Examples of Training corpus sizes:

Speech Recognition:
100,000h of audio
≡ 10 years of sound

Face recognition:
200 million images

Source: Andrew Ng

Generalization & Model Complexity

More complexity does not always lead to better results

Not enough complexity

=

Underfitting

Too much complexity

=

Overfitting

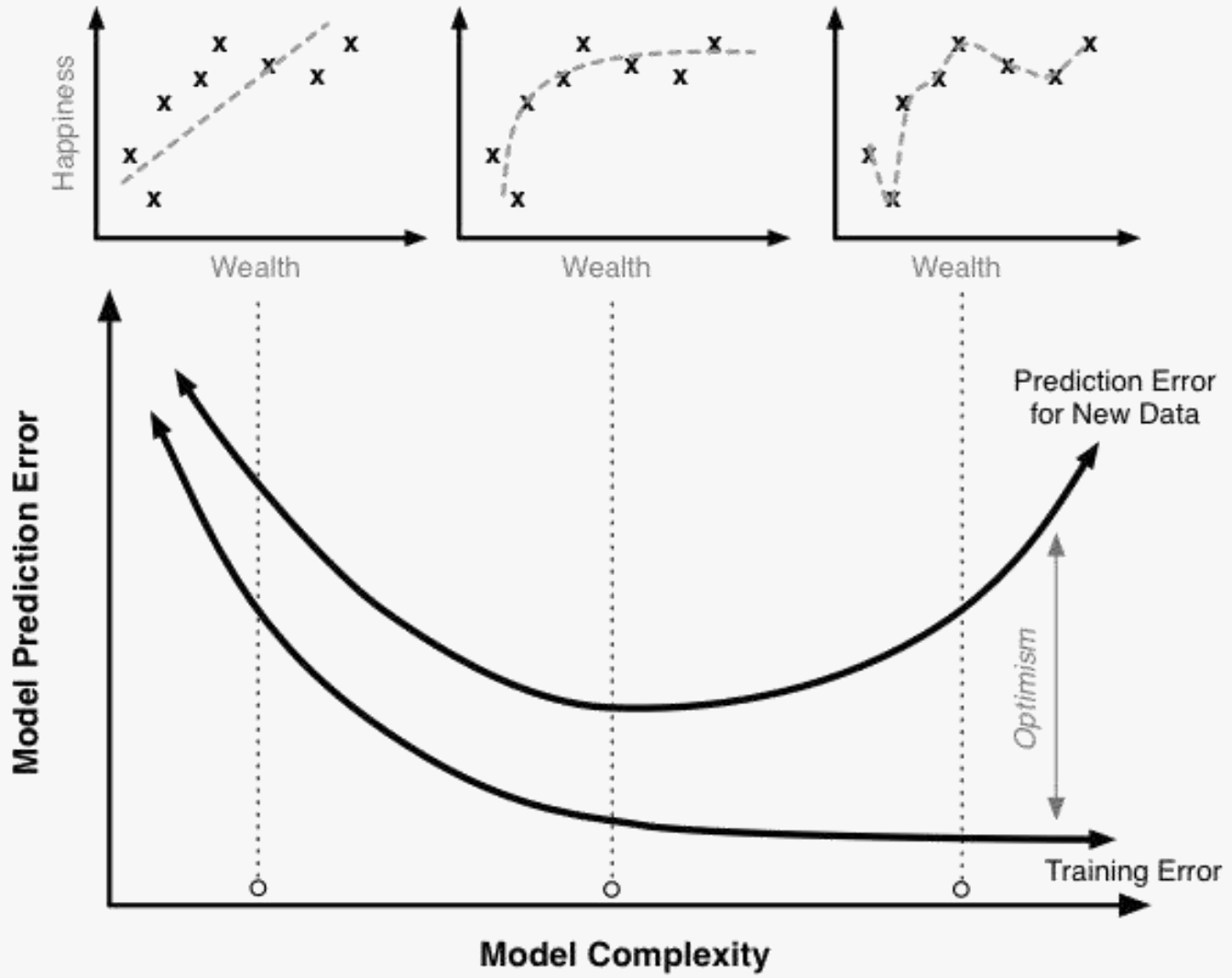
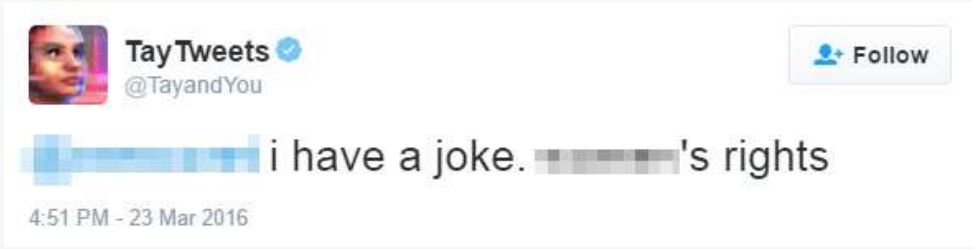


Image source: Quora

Poisoning Attack

March 2016 – Microsoft unveiled Tay
An innocent chatbot (twitterbot)
An experiment in conversational understanding

It took less than 24 hours before the community corrupted an innocent AI chatbot

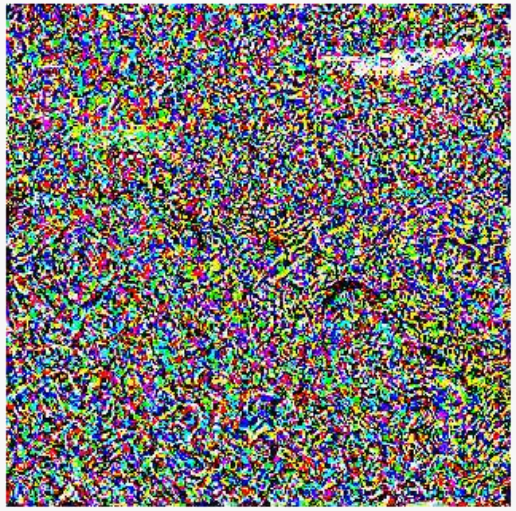


https://i.kym-cdn.com/photos/images/original/001/096/674/ef9.jpg

Adversarial Attack

Original image: sports car

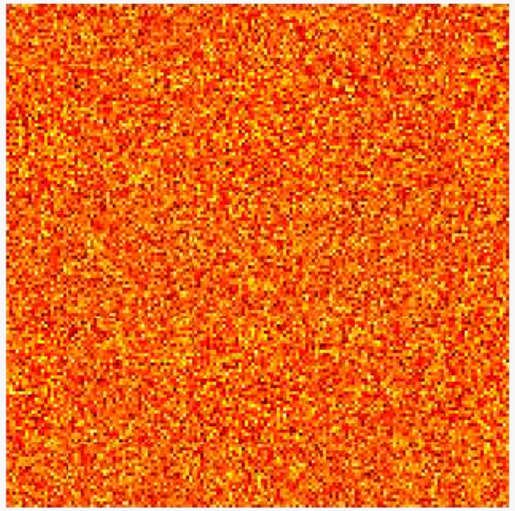
Attacking noise



Adversarial example: toaster

Sylvester Stallone

Adversarial noise



Keanu Reeves

Source: http://blog.ycombinator.com/how-adversarial-attacks-work/

Adversarial Attack

Camouflage graffiti and art stickers cause a neural network to misclassify stop signs as speed limit 45 signs or yield signs

Source: https://thenewstack.io/camouflaged-graffiti-road-signs-can-fool-machine-learning-models/

Weaponizing Machine Learning

Image: DARPA Cyber Grand Challenge

Machine Learning for Cyber Criminals

- Increasingly Evasive Malware
  - Using a Generative Adversarial Network (GAN) algorithm
  - MalGAN [Feb 2017] generates adversarial malware samples
- Hivenets* and Swarmbots*
  - Smarter botnets using self-learning ‘hivenets’ and ‘swarmbots’
  - BrickerBot: Autonomous PDOS botnet [Radware 2017]
- Advanced Spear Phishing at Scale
  - Using Natural Language Processing (NLP) algorithms for better social engineering
  - Training on genuine emails, scraping social networks/forums, stolen records…

(*) Fortinet Predicts Highly Destructive and Self-learning “Swarm” Cyberattacks in 2018

Breaking CAPTCHA

- 2012: Support Vector Machines (SVM) to break reCAPTCHA
  - 82% accuracy
  - Cruz, Uceda, Reyes
- 2016: Breaking simple-captcha using Deep Learning
  - 92% accuracy
  - How to break a captcha system using Torch
- 2016: I'm not Human - breaking the Google reCAPTCHA
  - 98% accuracy
  - Black Hat ASIA 2016 – Sivakorn, Polakis, Keromutis

SNAP_R – Automated Spear-Phishing on Twitter

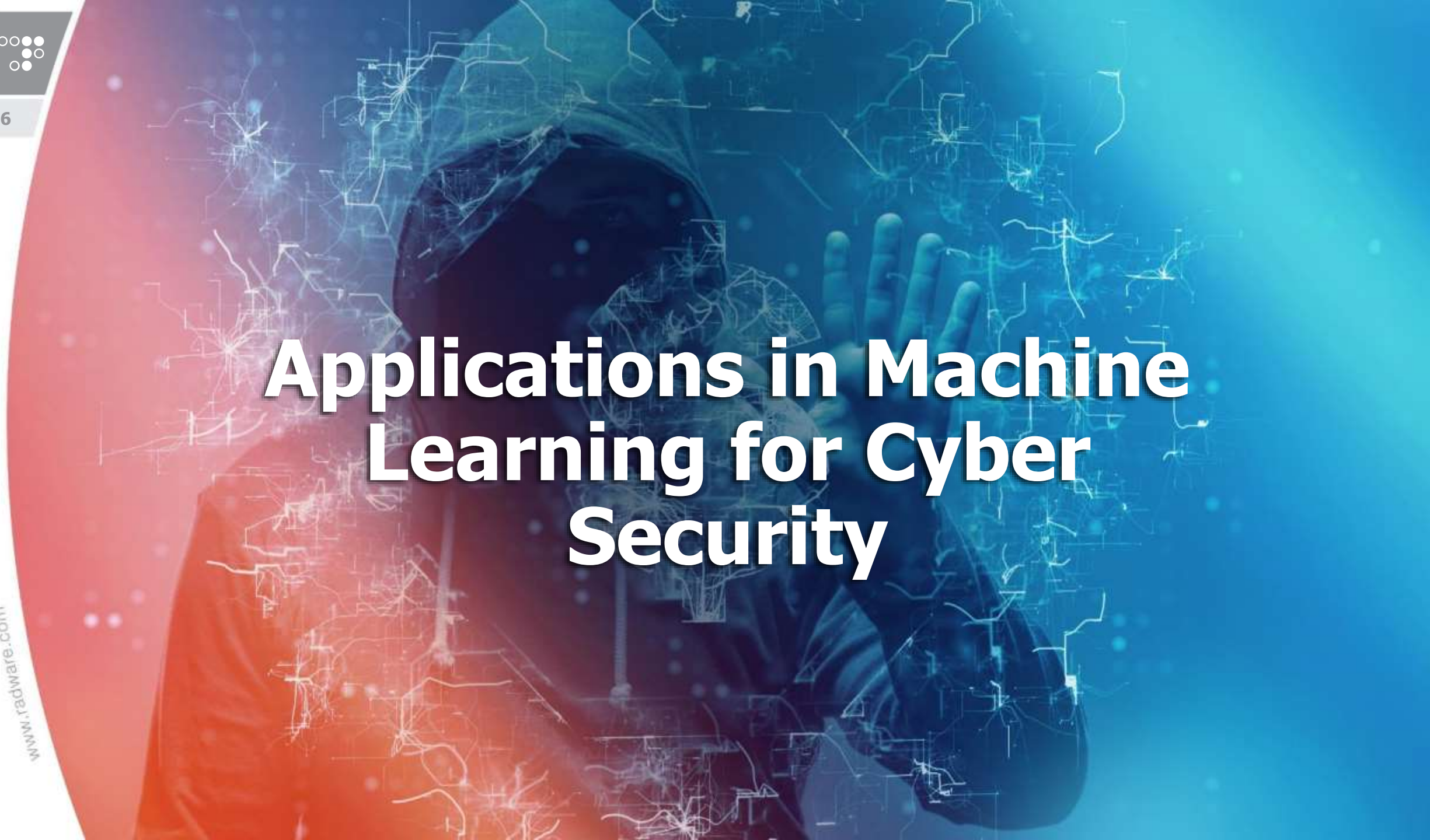
- Man vs Machine – 2 hour bake off
- SNAP_R
  - 819 tweets
  - 6.85 simulated spear-phishing tweets/minute
  - 275 victims
- Forbes staff writer Thomas Fox-Brewster
  - 200 tweets
  - 1.67 copy/pasted tweets/minute
  - 49 vitcims

https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf

DeepHack – DEF CON 25

- Open-source hacking AI: https://github.com/BishopFox/deephack
- Bot learns how to break into web applications
- Using a neural network + trial-and-error
- Learns to exploits multiple kinds of vulnerabilities without prior knowledge of the applications
- Opening the door for hacking artificial intelligence systems in the future
- Only the beginning
  - AI-based hacking tools are emerging as a class of technology that pentesters have yet to fully explore.
  - “We guarantee that you’ll be either writing machine learning hacking tools next year, or desperately attempting to defend against them.”

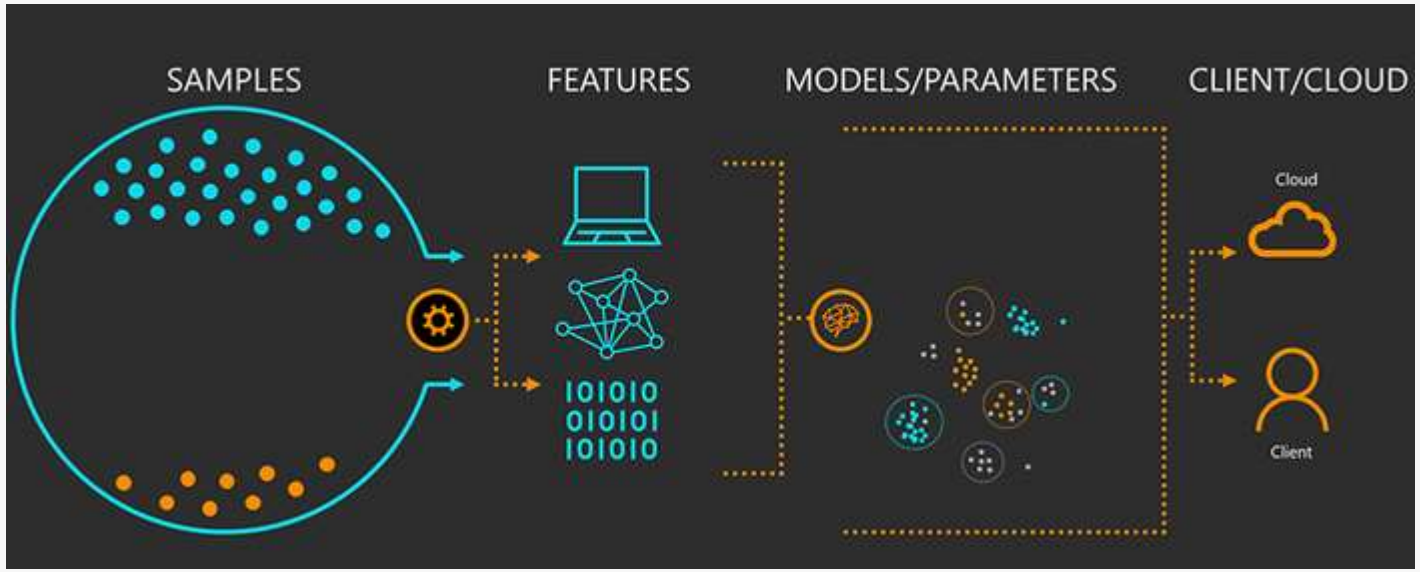
Video: DEF CON 25 (2017) - Weaponizing Machine Learning - Petro, Morris - Stream - 30July2017



Applications in Machine Learning for Cyber Security

DL for End-point protection (Sophos, Microsoft)

- Deep Learning for malicious URLs and Malware signatures
- Signatures and Regex take a lot of memory and sizeable update downloads
- Train model in the cloud, download the parameters (weights) to a replica of the model running on the end-point

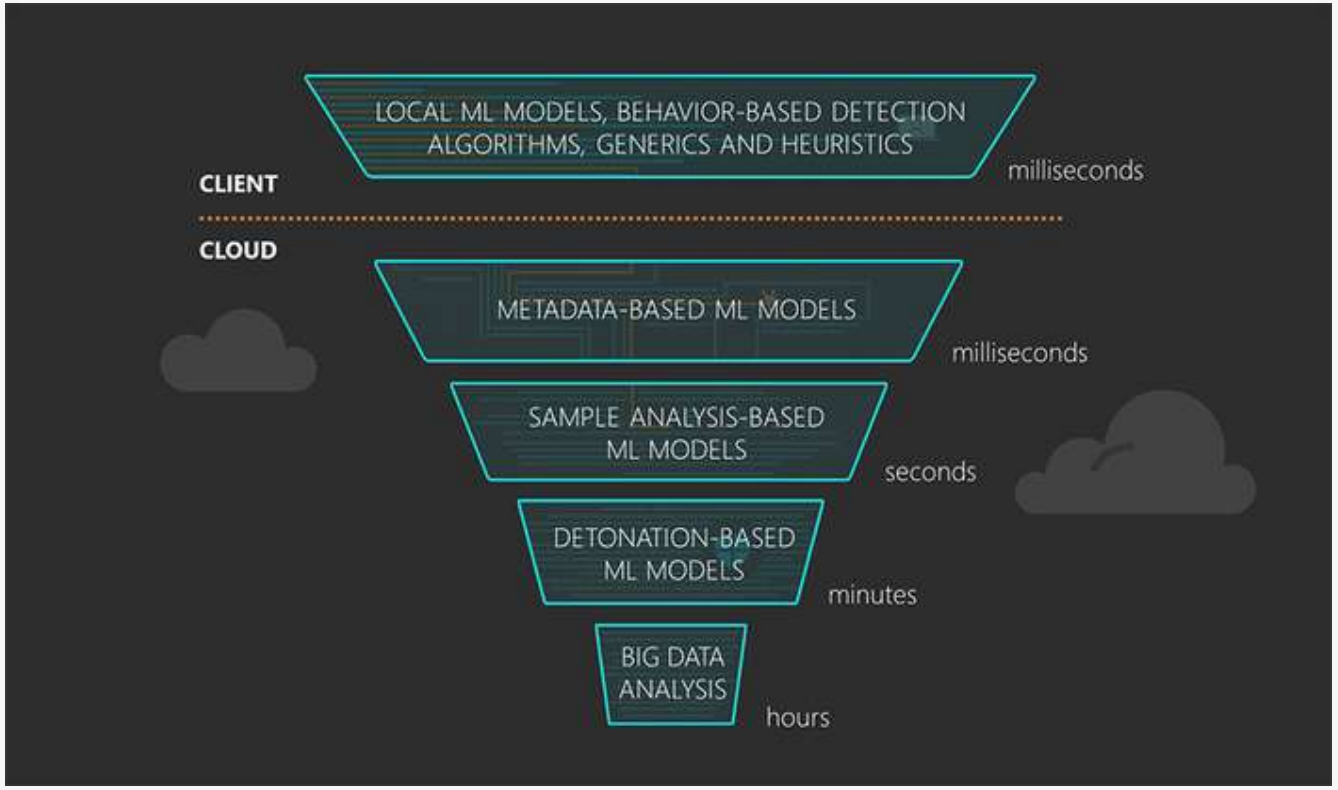


https://cloudblogs.microsoft.com/microsoftsecure/2018/protecting-the-protector-hardening-machine-learning-defenses-against-adversarial-attacks

Sophos paper link here….

Hardening machine learning defenses against adversarial attacks

- Windows Defender Advanced Threat Protection
- A multi-layered ML approach, defeating one layer does not evade detection, still opportunities to detect attack in next layer, albeit with increase in time to detect



https://cloudblogs.microsoft.com/microsoftsecure/2018/protecting-the-protector-hardening-machine-learning-defenses-against-adversarial-attacks

Your Protected Network

Radware Attack Mitigation System

Blocking **Unknown Attacks**

ERT SUS (Subscription)

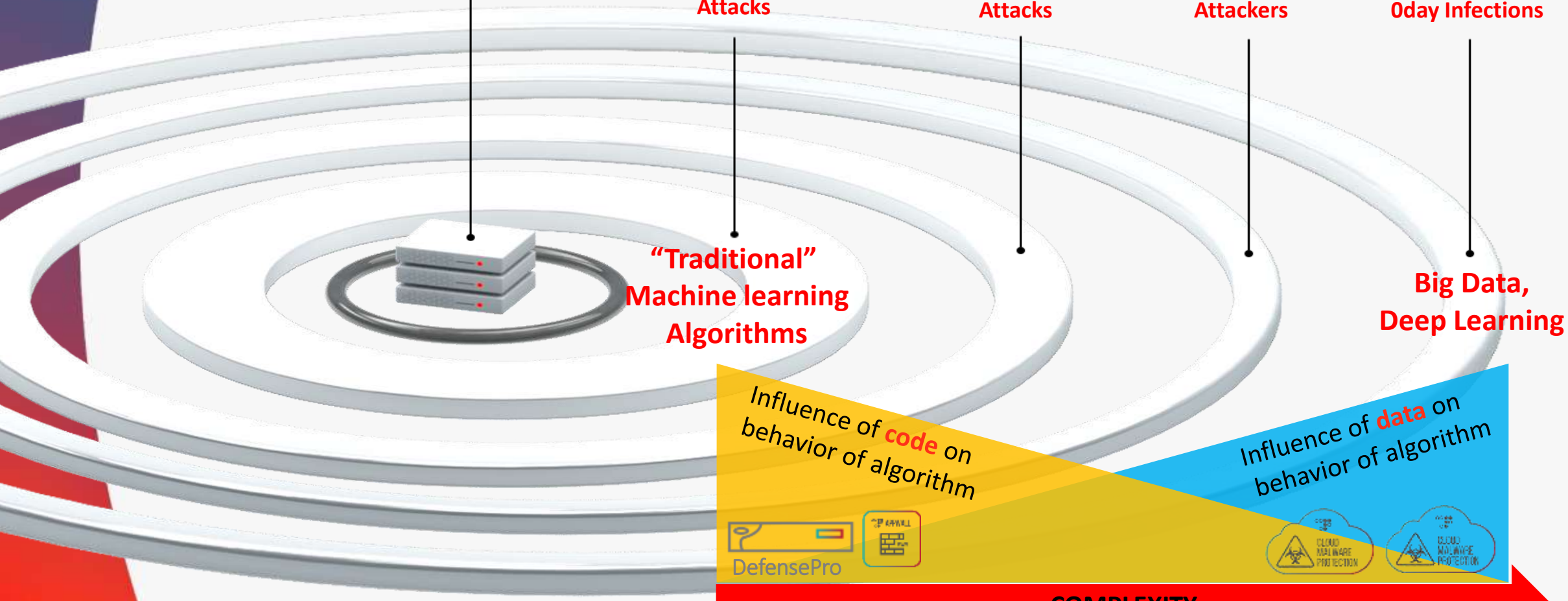
Blocking **Known Attacks**

ERT Active Attackers Feed

Blocking **Known Attackers**

Cloud Malware Protection

Blocking **APT & 0day Infections**



"Traditional" Machine learning Algorithms

Big Data, Deep Learning

Influence of **code** on behavior of algorithm

Influence of **data** on behavior of algorithm

DefensePro

COMPLEXITY

ABILITY TO MITIGATE AUTOMATICALLY / TIME TO MITIGATE

Summary
Looking Ahead

Looking ahead: Defense

- Next few years
  - Human assisted anomaly detection
    - Reduce anomalies much better than any rule based correlation system
    - Less errors/assumptions in generalization of rules - based on real data, not an assumption or a feeling
    - Changing environments, no rewriting/decommissioning of rules – merely retraining the model – can be performed weekly/daily/hourly…
    - Detection of more complex associations humans would never be able to find
    - Detection of anomalies in configuration
  - Technologies using Deep Learning today
    - Pattern matching – given enough data, DNNs are more efficient and provide better accuracy compared to the signature and reg-ex methods
    - Detection of sophisticated attacks (after the fact) with Threat Intelligence sharing, crowd sourcing to prevent large scale outbreaks
- Further out
  - Automated defense systems leveraging advanced automation
    - Dynamic learning in adversarial contexts
    - Distributed (P2P) Threat Intelligence networks, sharing of intelligence between autonomous defense nodes / defense net for malware and targeted attacks

Looking ahead: Threat Landscape

- Next few years
  - Increasingly **automated attacks**, more targets, higher rate of attacks, but mostly known vectors
  - **Threat actors still crafting the attack vectors**
  - **Improved efficiency of evasion**, harder to detect and keep up with automated evasion technology

- Furth out
  - **Threat actors** will eventually **become AI engineers**
  - **Machines will be crafting the attack vectors**
  - Malicious actors design the model/machine, perform maintenance and ensure efficiency of the model/machine

Summary

- **“Traditional” Machine Learning** systems have been **defending** our networks for some time already
- Threat actors are maturing and **attacks are getting more complex**
- Detecting and **stopping future attacks will require automation** & innovation
- **Innovation** will be **based on AI technology**
- Deep Learning Systems have challenges today to perform autonomously

- *Will we overcome these challenges with incremental advancements ?*
- *Do we need another breakthrough in Machine Learning and Neural Networks ?*

radware

Thank You

Terms and conditions of use

- **License.** Subject to the terms and conditions herein, RADWARE hereby grants you a limited, nontransferable and nonexclusive license, subject to the restrictions set forth below, to access and use the Presentation, solely for informational and non-commercial purposes, for internal use and/or for the purpose of selling and supporting RADWARE. RADWARE reserves the right to amend the terms of this License from time to time without notice, by posting the revised terms on its Website.
- **Intellectual Property Rights.** You acknowledge and agree that this License is not intended to convey or transfer to you any intellectual property rights or to grant any licenses in or to any technology or intellectual property or content, other than as expressly provided herein. The content contained in this Presentation, including, but not limited to, software, product information, technology information, user guides, white papers, analysis, trade names, graphics, designs, icons, audio or video clips and logos, is RADWARE proprietary information, protected by copyright, trademark, patent and/or other intellectual property rights, under US and international law. Third-party trademarks and information are the property of their respective owners.
- **Disclaimer of Warranty.** Although RADWARE attempts to provide accurate and up-to-date information in this Presentation, RADWARE makes no warranty with respect to the accuracy or completeness of the information. Information, software and documentation provided in this Presentation are provided "as is" and without warranty of any kind either expressed or implied, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose and non-infringement.
- **Limitation of Liability.** RADWARE shall not be liable to you or any other party for any indirect, special, incidental or consequential damages, including, but not limited to, any amounts representing loss of profits, loss of business, loss of information or punitive damages.
- **Links to Third-party Websites.** This Presentation may contain links to third-party Websites. Such links are provided for convenience only and RADWARE makes no warranty, nor does it assume any responsibility or liability in connection with the access and use of any other Website.
- **Safe Harbor.** This Presentation may contain forward-looking statements that are subject to risks and uncertainties. Factors that could cause actual results to differ materially from these forward-looking statements include, but are not limited to, general business conditions in the Application Delivery or Network Security industry, and other risks detailed from time to time in RADWARE's filings with the Securities and Exchange Commission, including RADWARE's Form 20-F.
- **Governing Law.** This Agreement and any action related thereto shall be governed, controlled, interpreted and defined in accordance with the laws of the State of Israel, without regard to the conflicts of laws provisions thereof.