

# RSA® Conference 2018

San Francisco | April 16–20 | Moscone Center

SESSION ID: SEM-M04

## DATA SCIENCE— HOW IT'S BECOMING A PIVOTAL PART OF IT SECURITY, AUTOMATION & ORCHESTRATION

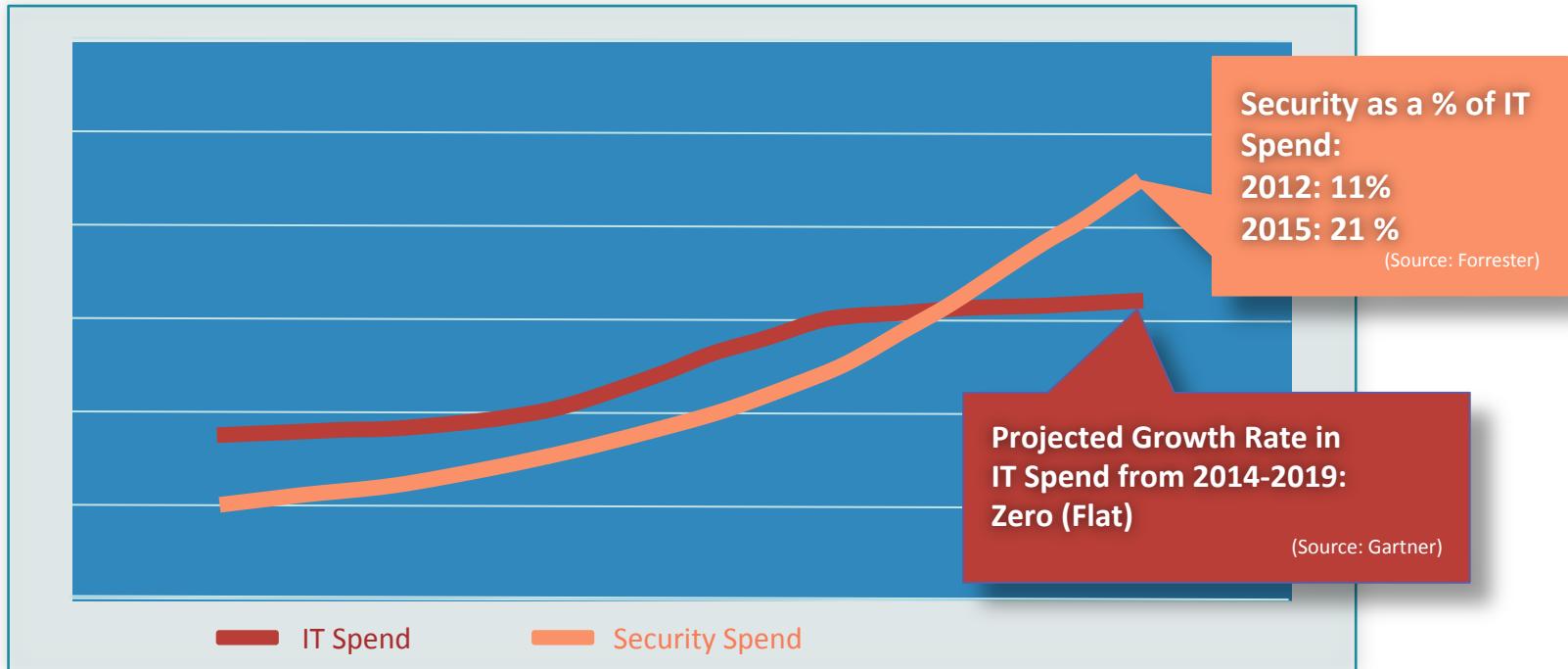
**Vijay Ganti**

Head of Product & ML/AI Research  
Security Products Group  
VMware





# Security spend: outpacing IT spend 2:1



Source: IDC

© 2018 VMware Inc. All rights reserved.

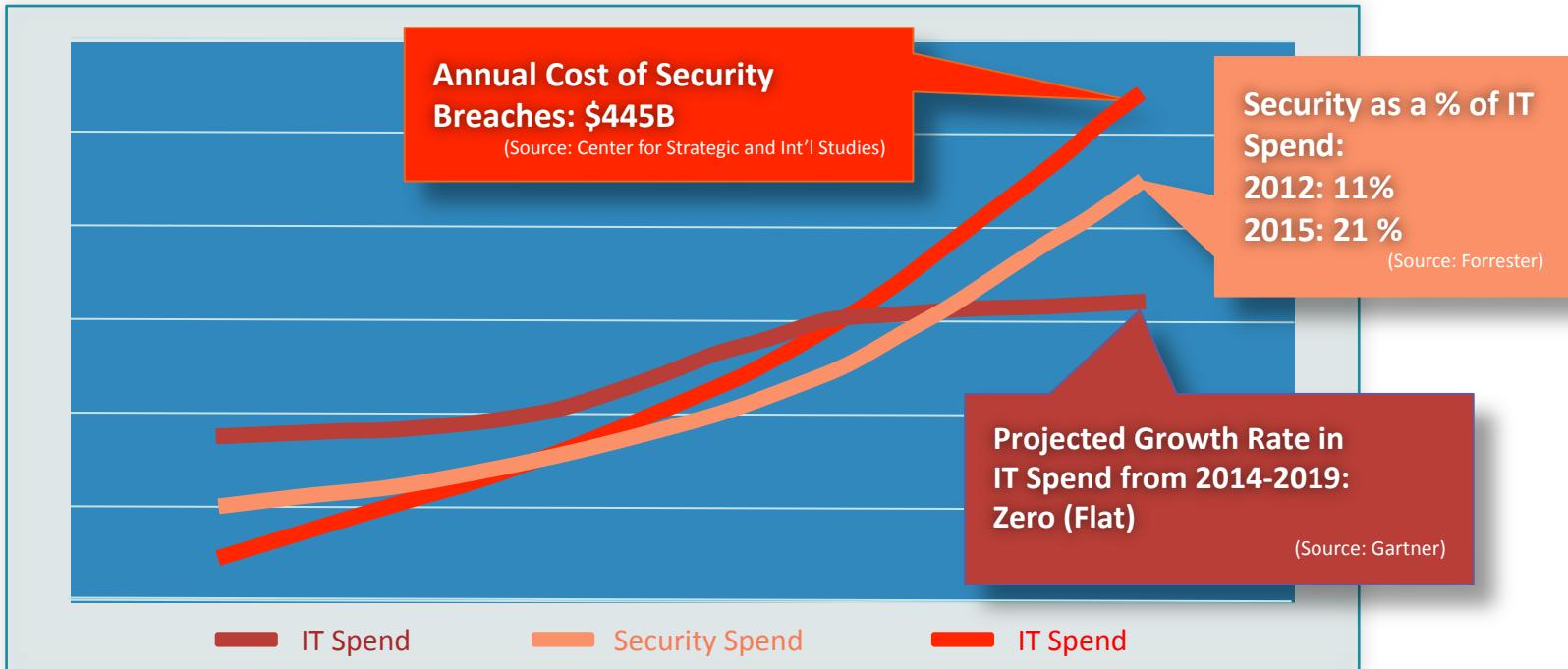
RSA Conference 2018

# Information Security Spending >\$80B in 2016\*



#RSAC

# A Picture of Diminishing Returns





## Infiltration

Attack vector/malware  
Delivery mechanism  
Entry point compromise



## Propagation

Escalate privileges  
Install C2\* infrastructure  
Lateral movement



## Extraction

Break into data stores  
Network eavesdropping  
App-level extraction



## Exfiltration

Parcel and obfuscate  
Exfiltration  
Cleanup



We over invest in Infiltration Prevention

We under invest in Resilience



## Propagation

- Escalate privileges
- Install C2\* infrastructure
- Lateral movement



## Extraction

- Break into data stores
- Network eavesdropping
- App-level extraction



## Exfiltration

- Parcel and obfuscate
- Exfiltration
- Cleanup

# We're Watching The Wrong Thing

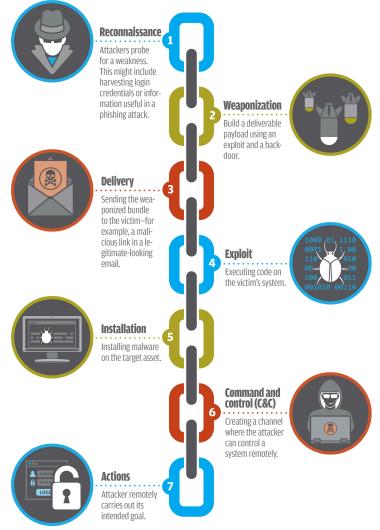


# NOT LEVERAGING ASYMMETRY



## ASYMMETRY

What is the **CYBER KILL CHAIN?**  
The cyber kill chain, created by Lockheed Martin, describes the phases or stages of a targeted attack. Each stage presents an opportunity to detect and react to an attack.



## HOME COURT ADVANTAGE

RSA Conference 2018



## Infiltration

Attack vector/malware  
Delivery mechanism  
Entry point compromise



We over invest  
in  
Chasing Threats



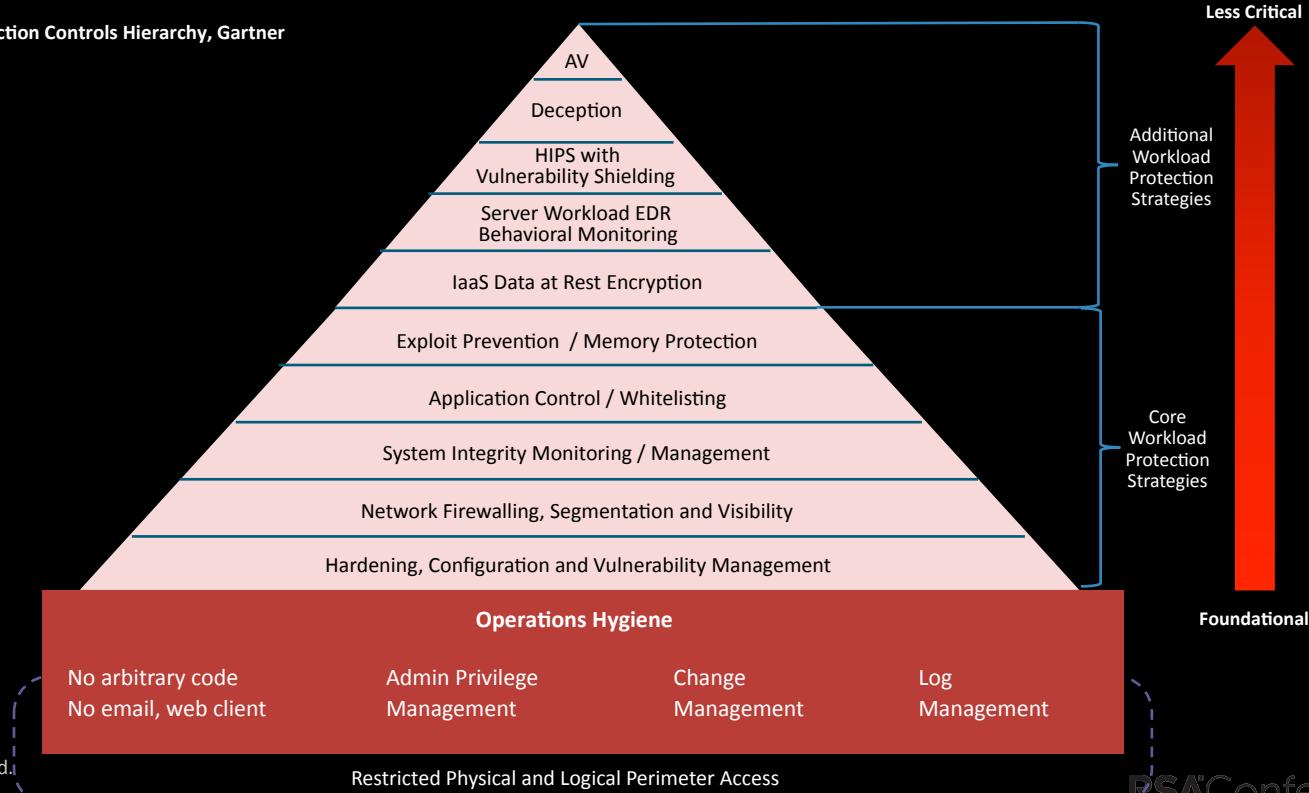
We under invest  
in Shrinking  
the Attack Surface

# Focus on shrinking the attack surface

#RSAC

## Shrinking the Attack Surface

Figure 1. Cloud Workload Protection Controls Hierarchy, Gartner





## Cyber Threats

Residual Risk

#RSAC



Micro-Segmentation



Least Privilege



Encryption



Multi-Factor Authentication



Patching

## Cyber Hygiene

Attack Surface



## Infiltration

Attack vector/malware  
Delivery mechanism  
Entry point compromise

We over invest  
in  
Chasing Threats

We under invest  
in Shrinking  
the Attack Surface

We Align Security  
To  
Infrastructure

Rather than Align  
To  
Applications & Data



# The Architectural Gap



# Rethinking CYBER Security



## Cyber Threats

Residual Risk



## Integrated Ecosystem



## Securable Infrastructure



## Cyber Hygiene

Attack Surface

# CHASING BAD...

CHASING  
BAD

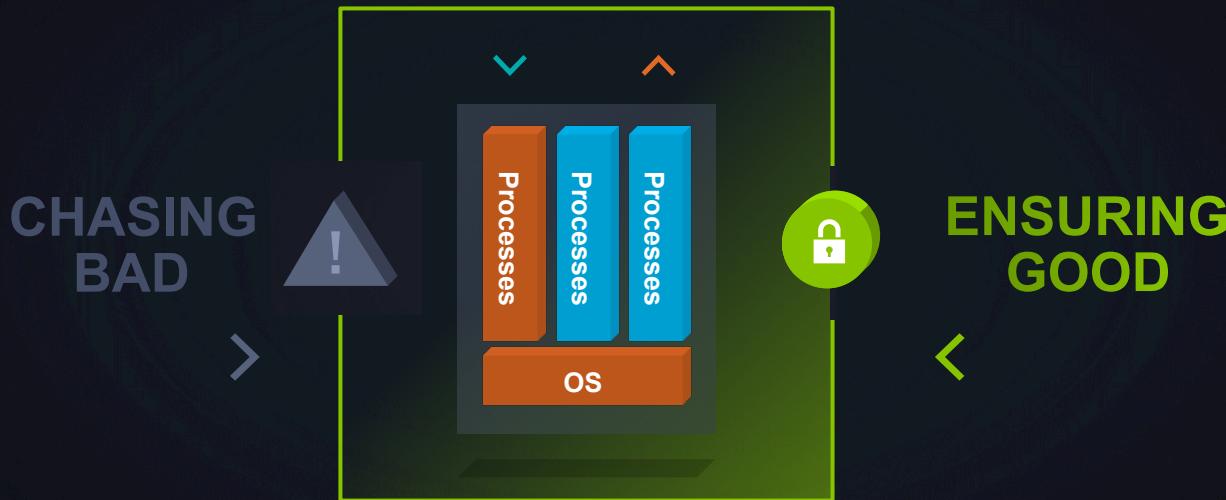


```
001010010100101001010101010  
1010100101010101010101010  
11111010101001010010100000MAL  
WARE010100101001010010101010  
0101001010010101010101010  
010101010101101010111101  
010100101010000010101001  
0101001001010010101001010  
01010010010101010101001010  
1010101001010101010101010  
010100101010100101010100  
1010010100101010101010100  
10101010101100101010101111010  
101001010010100000MALWARE01  
0100100101001001010100101001
```



The diagram illustrates a memory hierarchy. At the bottom, the word "os" is written vertically. Above it, the word "Processes" is also written vertically. Three arrows point upwards from "os" to "Processes": one blue arrow pointing to the first "Processes" block, one red arrow pointing to the second, and one green arrow pointing to the third. The text above "Processes" consists of binary code. The word "WARE" appears in the middle of the binary sequence, followed by the word "MALWARE" at the very top. The entire sequence is enclosed in a light gray rectangular box with a thin black border.

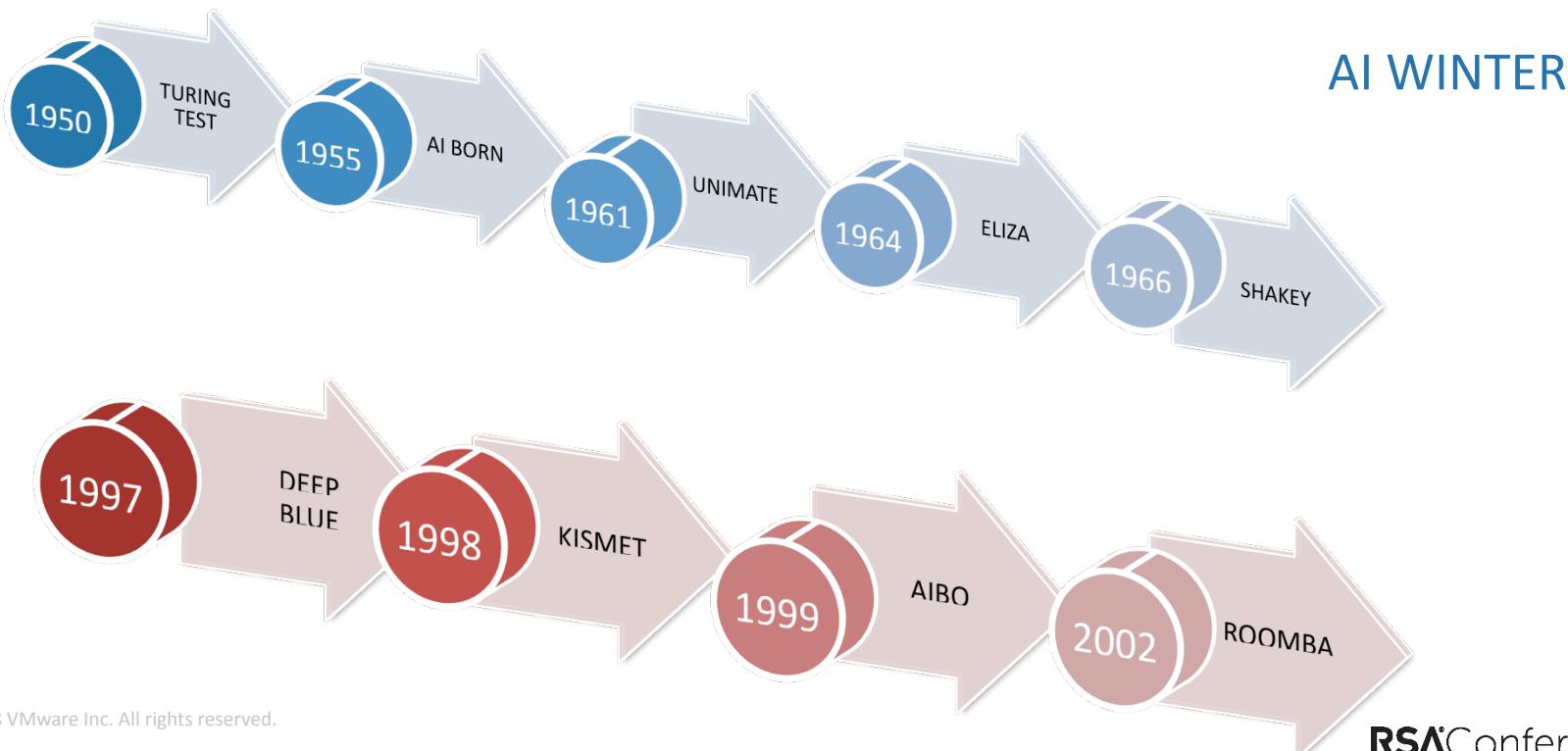
# CHASING BAD... VERSUS ENSURING GOOD



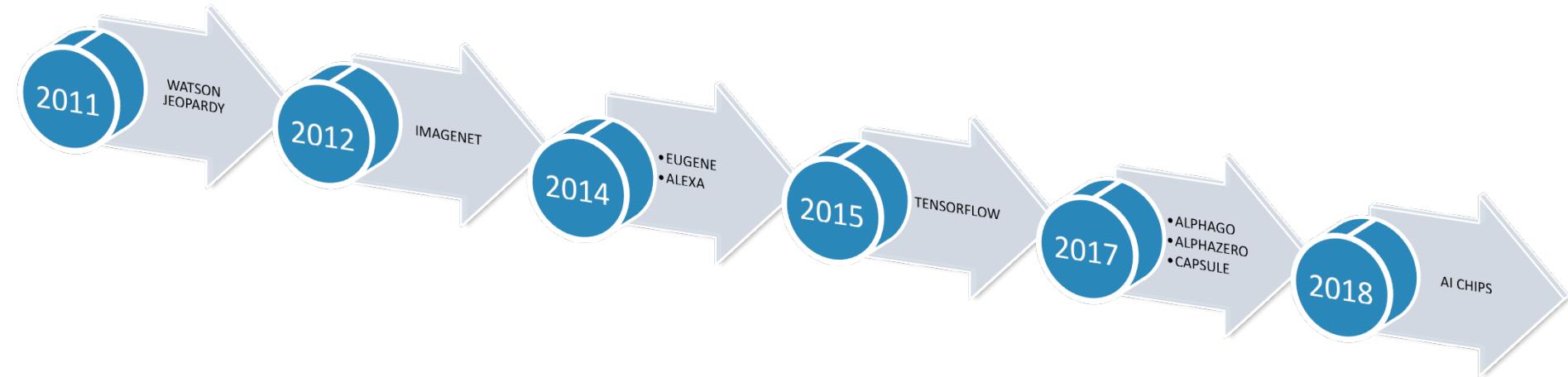
# How can AI help with, Threat detection, cyber hygiene and modern security operations?



# A brief history of AI

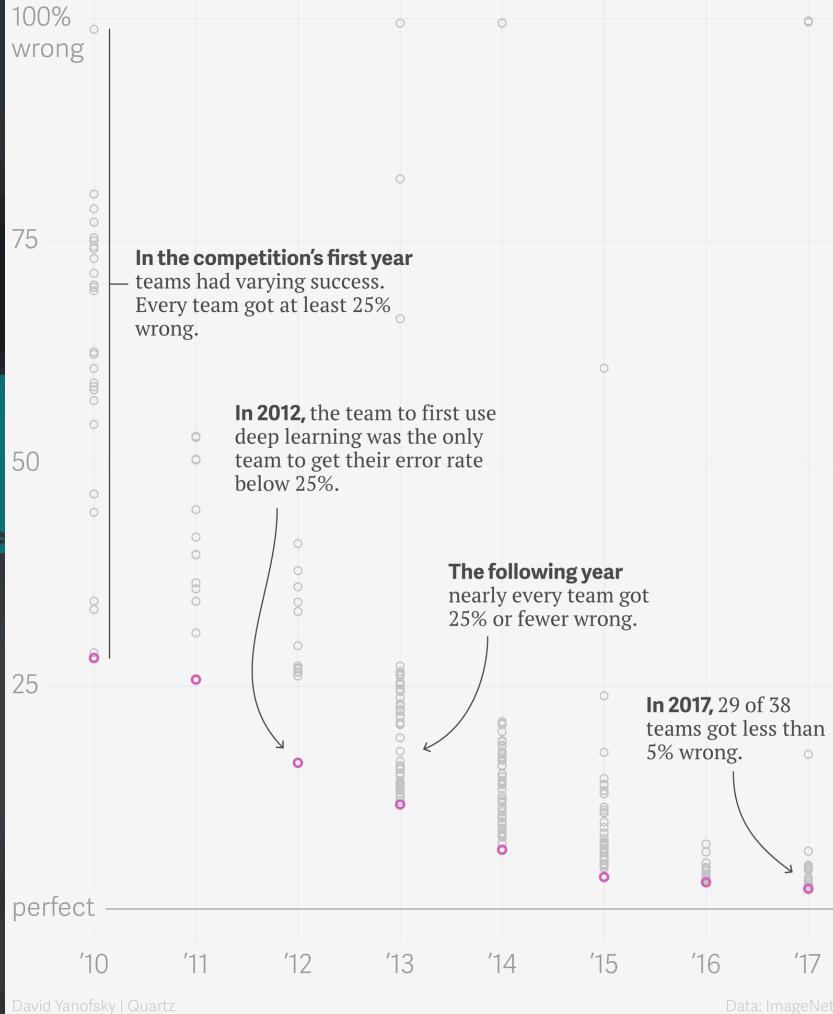


# A brief history of AI



# 2012

## ImageNet Large Scale Visual Recognition Challenge results



# RECENT AI ACCOMPLISHMENTS



JANUARY 25, 2017

## Deep learning algorithm does as well as dermatologists in identifying skin cancer

*In hopes of creating better access to medical care, Stanford researchers have trained an algorithm to diagnose skin cancer.*

[www.theguardian.com](http://www.theguardian.com)

**AlphaZero AI beats champion chess program after teaching itself in four hours**

## Microsoft's AI beats Ms. Pac-Man

Posted Jun 15, 2017 by Brian Heater (@bheater)

Hybrid Reward Architecture



Level: 201

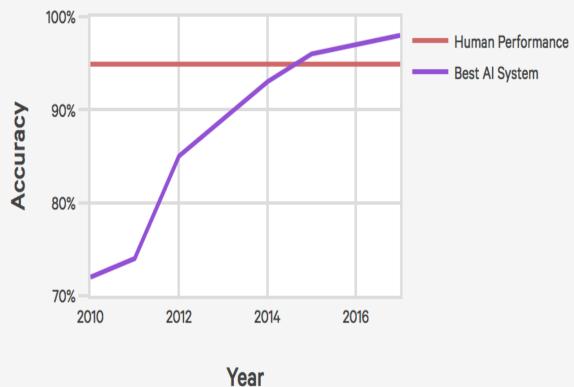
 Maluuba  
A Microsoft company

-	30435	x 10 =	304350
■	801	x 50 =	40050
■■	17	x 200 =	3400
■■■	6	x 400 =	2400
■■■■	3	x 800 =	2400
■■■■■	1	x 1600 =	1600
●	42	x 100 =	4200
●●	40	x 200 =	8000
●●●	33	x 500 =	16500
●●●●	43	x 700 =	30100
●●●●●	48	x 1000 =	48000
●●●●●●	47	x 2000 =	94000
●●●●●●●	89	x 5000 =	445000
			1000000

# RECENT AI ACCOMPLISHMENTS WITH TECHNICAL PERFORMANCE



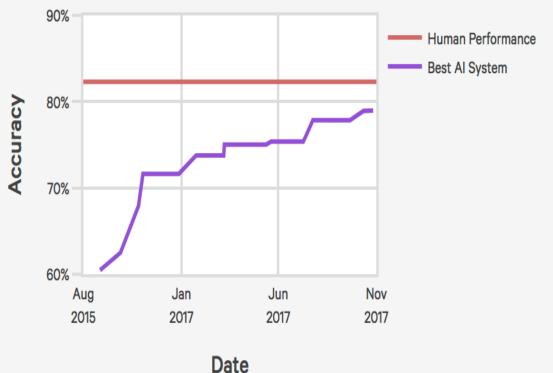
Object Detection, LSVRC Competition



Source: image-net.org

AIINDEX.ORG

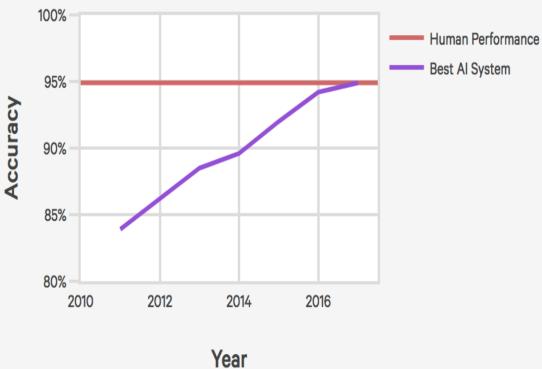
Question Answering, SQuAD v1.1



Source: stanford-qa.com

AIINDEX.ORG

Speech Recognition, Switchboard HUB5'00



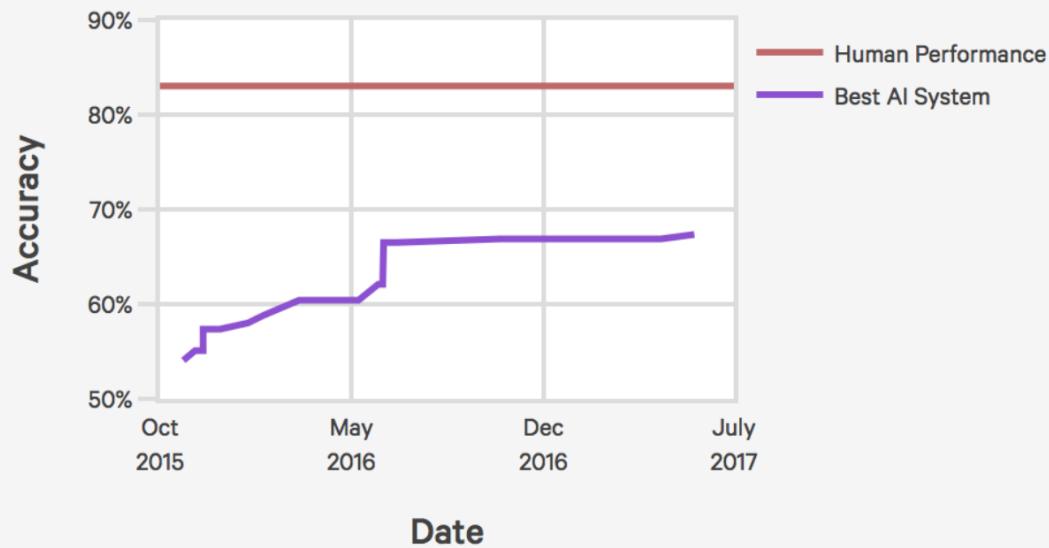
Source: Electronic Frontier Foundation, AI Progress Metrics

AIINDEX.ORG

# Long way to go before generalized AI



## Visual Question Answering, VQA 1.0



# What changed in AI since last time round ?



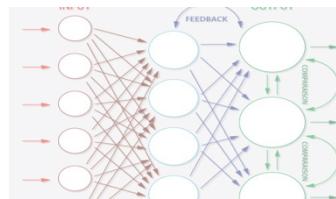
Lots of Data



Lots of Compute



Use cases /w lots of  
money



Algorithmic  
Innovation

# Challenges in Applying AI/ML to chasing the bad guys



## A PERFECT STORM

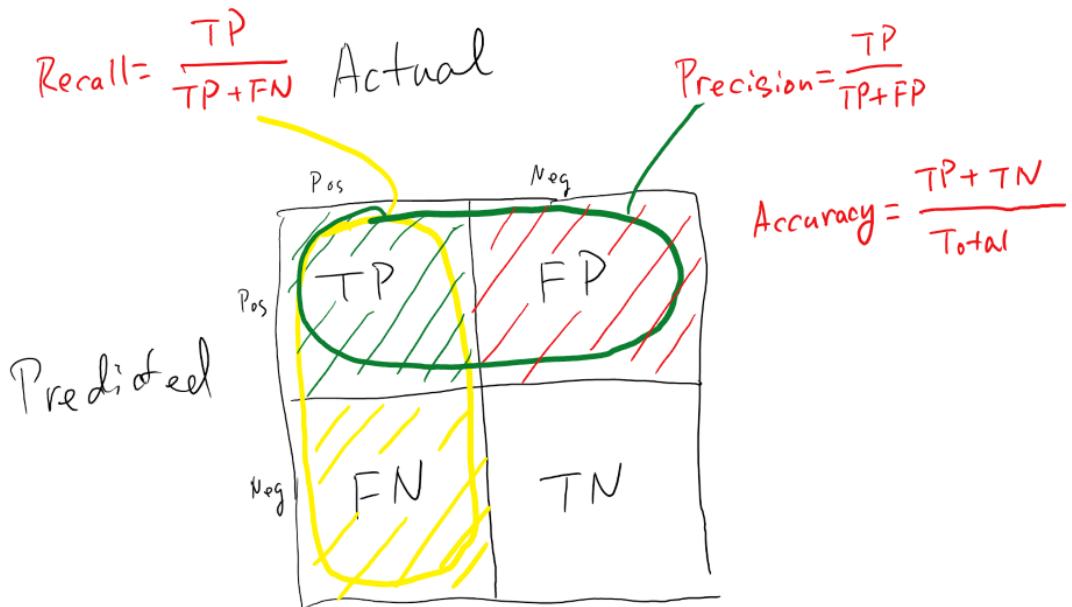
Adversarial

No Rules

Data Scarcity

**CHASING BAD IS HARD FOR EVEN AI/ML**

# False positives vs false negatives - Cost



- Huge focus on false negatives
  - Is the software able to detect attacks or malware?
- Little focus on false positives
  - Is your software throwing alerts when it should not ?
  - 10,000 to 150,000 alerts per day.

# Deep Learning vs Traditional Machine Learning



- ❑ What is the interpretability of the predictions?
- ❑ Performance – Accuracy (Precision & Recall)
- ❑ Amount of data needed for training models
- ❑ Where does the domain expertise go? (feature engineering vs designing networks)

# What About ensuring Good?



AI is a must to help achieve cyber hygiene in dynamic environments, at scale

---

- Understand Intended Application Composition
- Understand Intended Application Behavior
- Understand Intended Operational Changes

**THIS IS NOT ADVERSARIAL, HAS RULES & THERE IS DATA**

---

# Understanding application composition



- ❖ Service/Application Identification or Discovery
- ❖ Distributed Application Component Identification
- ❖ VM Composition

# Understanding application Behavior



- ❖ Coarse Grain Network Behavior
- ❖ Fine Grain Network Behavior
- ❖ System Calls
- ❖ File System Organization

# Understanding OPERATIONAL processes and CHANGES



- ❖ Routine application changes (upgrades/patches)
- ❖ Routine infrastructure changes
- ❖ Operational activities (backup/restore, monitoring, troubleshooting, administering)

# What more can AI do?



AI should be the fabric of your security operations

---

- Empower security analysts with contextual information for quick resolution
  - Use analysis bots for reducing burden on security analysts with deeper, targeted, automated and non-disruptive analysis in response to high-level anomalies
-

# Takeaways



AI should be the fabric of your security operations

---

- AI is eating the world but it's not likely to be the silver bullet for threat detection
  - Cyber hygiene is foundation for cybersecurity & AI will transform cyber hygiene at scale
  - AI will make security analysts super human
-



# Thank you

VMware

@vijayganti

Head of Product and ML/AI Research