Booz | Allen | Hamilton

MACHINE LEARNING IN CYBER SECURITY

Fact, Fantasy, and Moving Forward

DECEMBER 2018

CONSULTING | ANALYTICS | DIGITAL SOLUTIONS | ENGINEERING | CYBER

INTRO

Dan Liebermann
Commercial Advanced Analytics
Booz Allen Hamilton

Fact and Fantasy

IT IS OFTEN SAID THAT CYBER SECURITY SOLUTIONS MUST EVOLVE

ADVANCED ATTACKERS AND TECHNIQUES

Attackers are more sophisticated than ever before and are adopting AI and analytics tools that are becoming increasingly easy to access

CYBER CRIME FOR SALE

From ransomware to DDoS attacks, professional cyber criminals are selling commoditized hacking as a service and victimizing more customers at a never-before-seen scale

GROWING COLLATERAL DAMAGE

Nation state cyber operations extend to affect non-state targets as malicious software replicates itself well beyond the targeted network and disruptions ripple through your supply chain

MACHINE LEARNING IS OFTEN DISCUSSED AS A KEY PART OF THAT EVOLUTION

MORE DATA AND COMPUTE

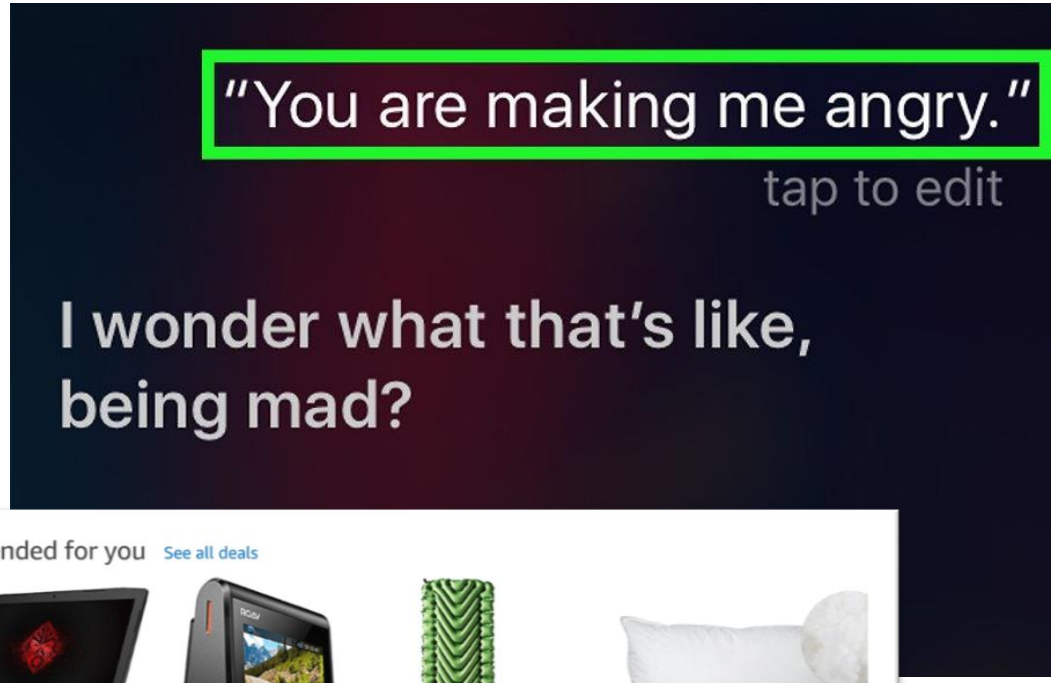
- *Explosion in Computing Power*
- *Exponential Data Volume Growth*
- *Variety of Data Sources and Formats*
- *Data Collected at Faster Velocity*

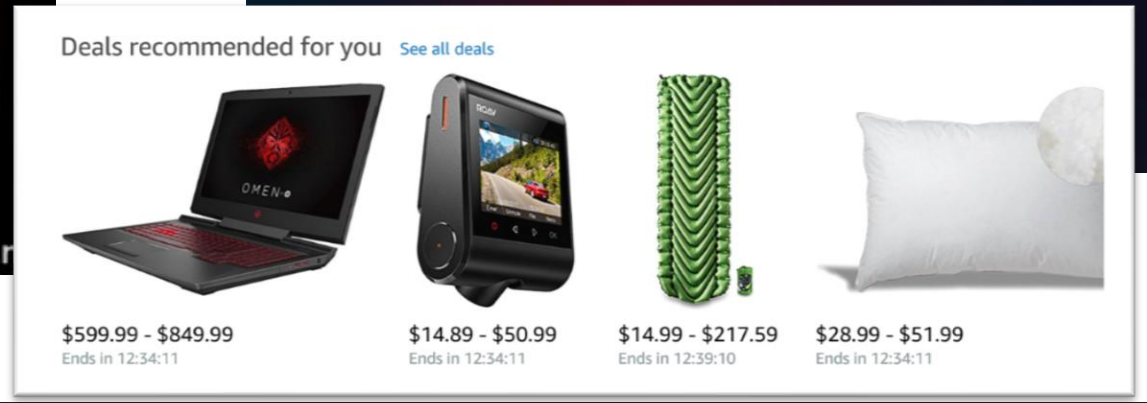
More opportunities than ever to use Machine Learning than ever before

AT A LOWER COST THAN EVER BEFORE

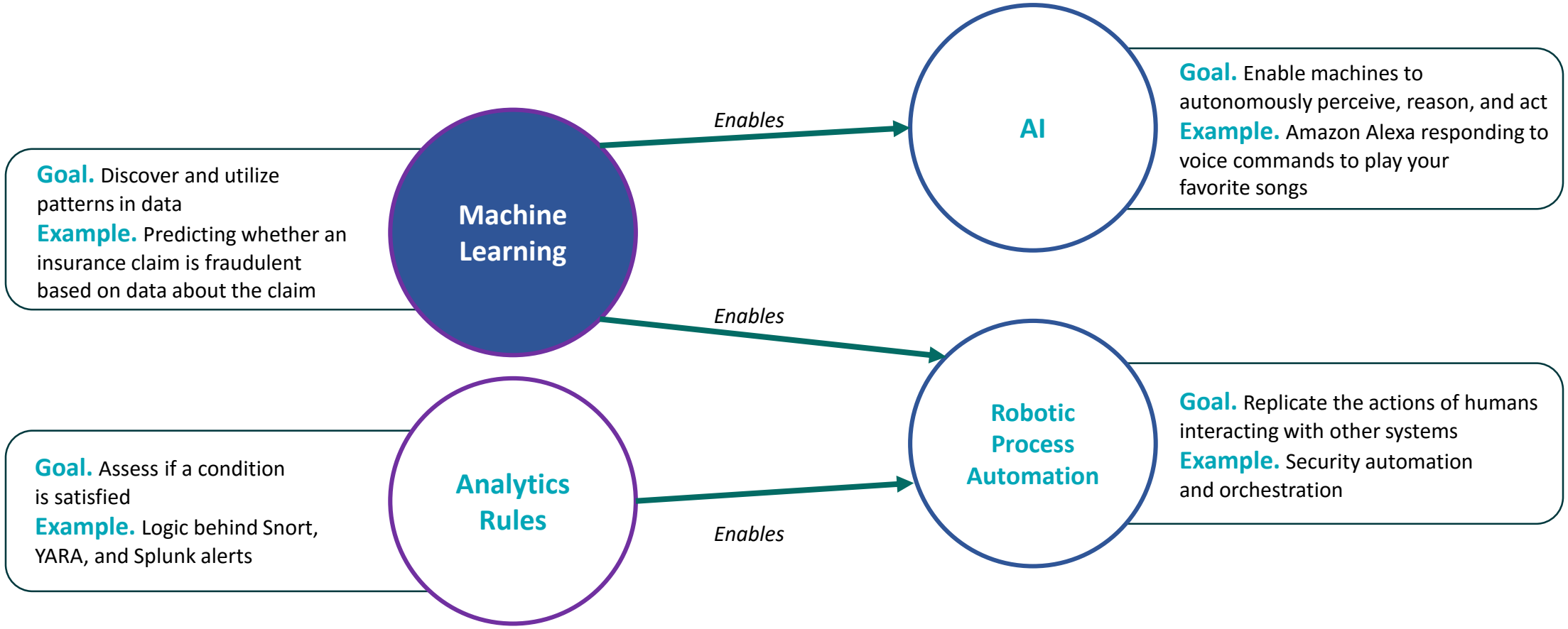
- *Lower Cost of Computing*
- *Affordable Cloud Infrastructure*
- *Free Open-Source Tools*
- *Community Code Sharing*

WHAT MIGHT COME TO MIND WHEN YOU HEAR THE TERM MACHINE LEARNING?





LET'S START BY CLARIFYING SOME TERMS



VENDORS HAVE FLOODED THE MARKET WITH CLAIMS THAT ARE TOO GOOD TO BE TRUE

"High accuracy, no noise"

"Uses machine learning and data science so anyone can get the same results as an expert in seconds"

"No endpoint protected by our product has EVER been breached"

"29x better productivity"

"Just like a veteran security expert"

"Impossible to deceive, unlike pre-canned algorithmic processes used by other security tools"

"No update ever needed"

"Automatically detects and classifies everything"

WHY PURSUE MACHINE LEARNING OR ANALYTICS AT ALL?

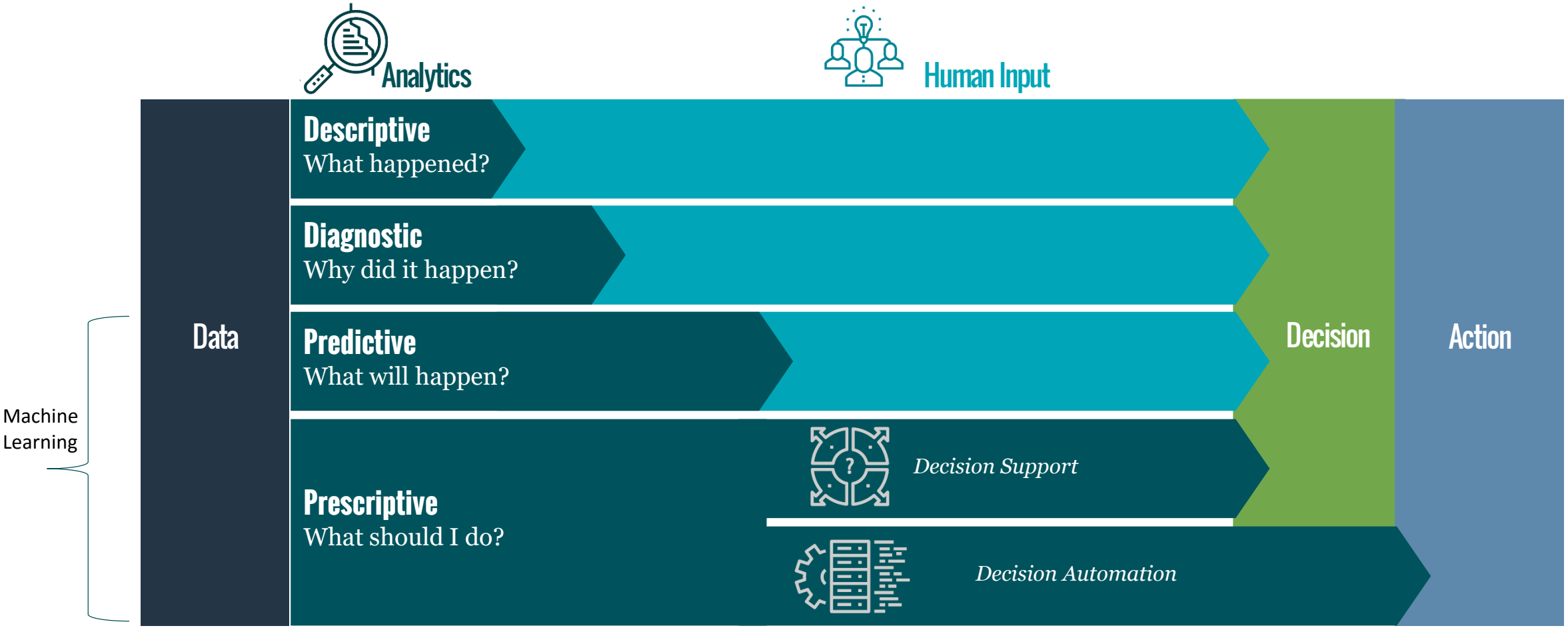
*The goal is to get **better intelligence**…*

*…better **risk modelling** and **prediction***

*…better **prioritization***

*…better **classification***

MACHINE LEARNING CONVERTS DATA INTO DECISIONS AND ACTIONS BETTER AND FASTER



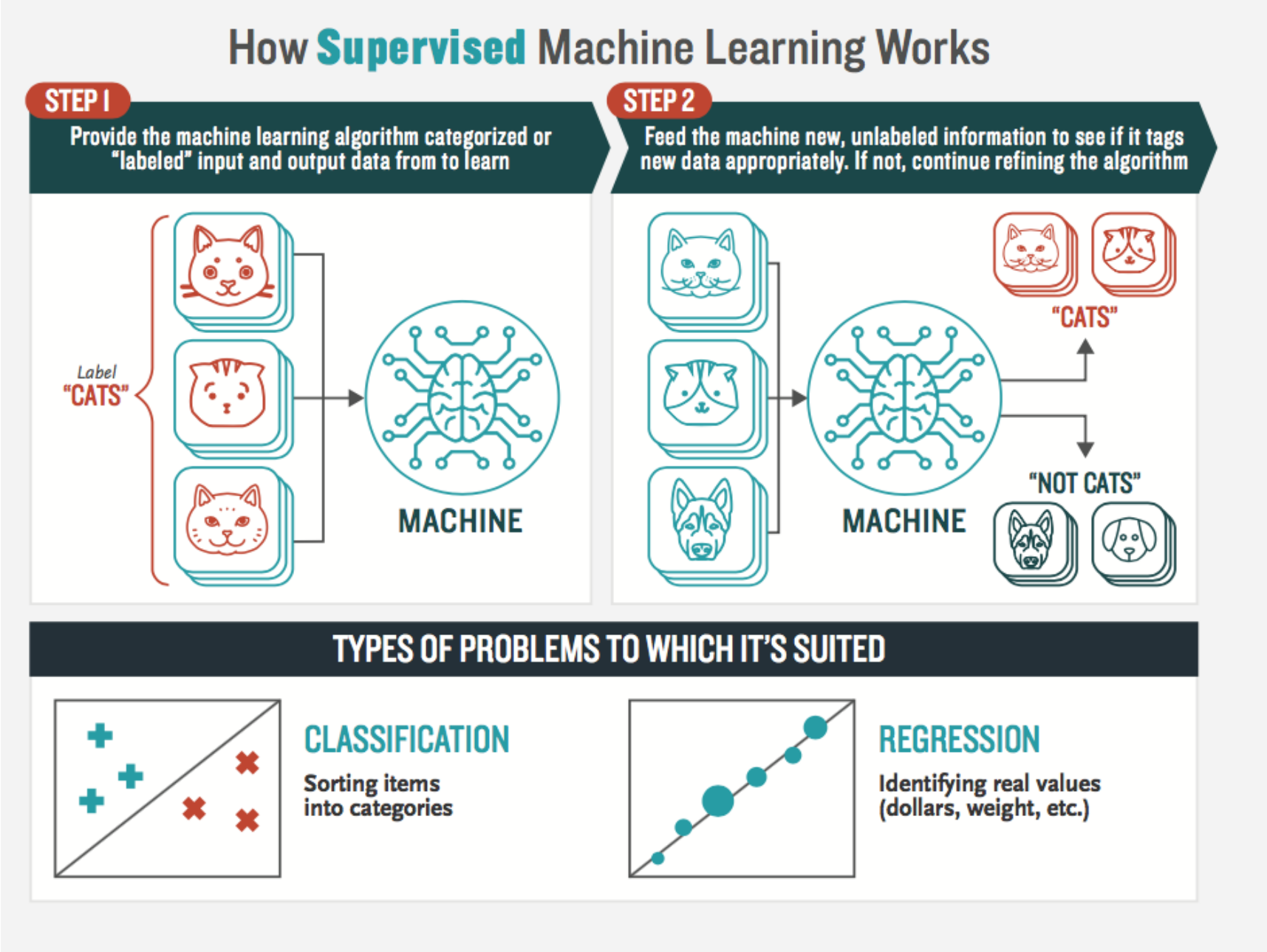
SIMPLY DESCRIBED, THERE ARE TWO COMMON MACHINE LEARNING APPROACHES

Supervised Machine Learning
Uses a labeled set (“I know what bad behavior looks like”)

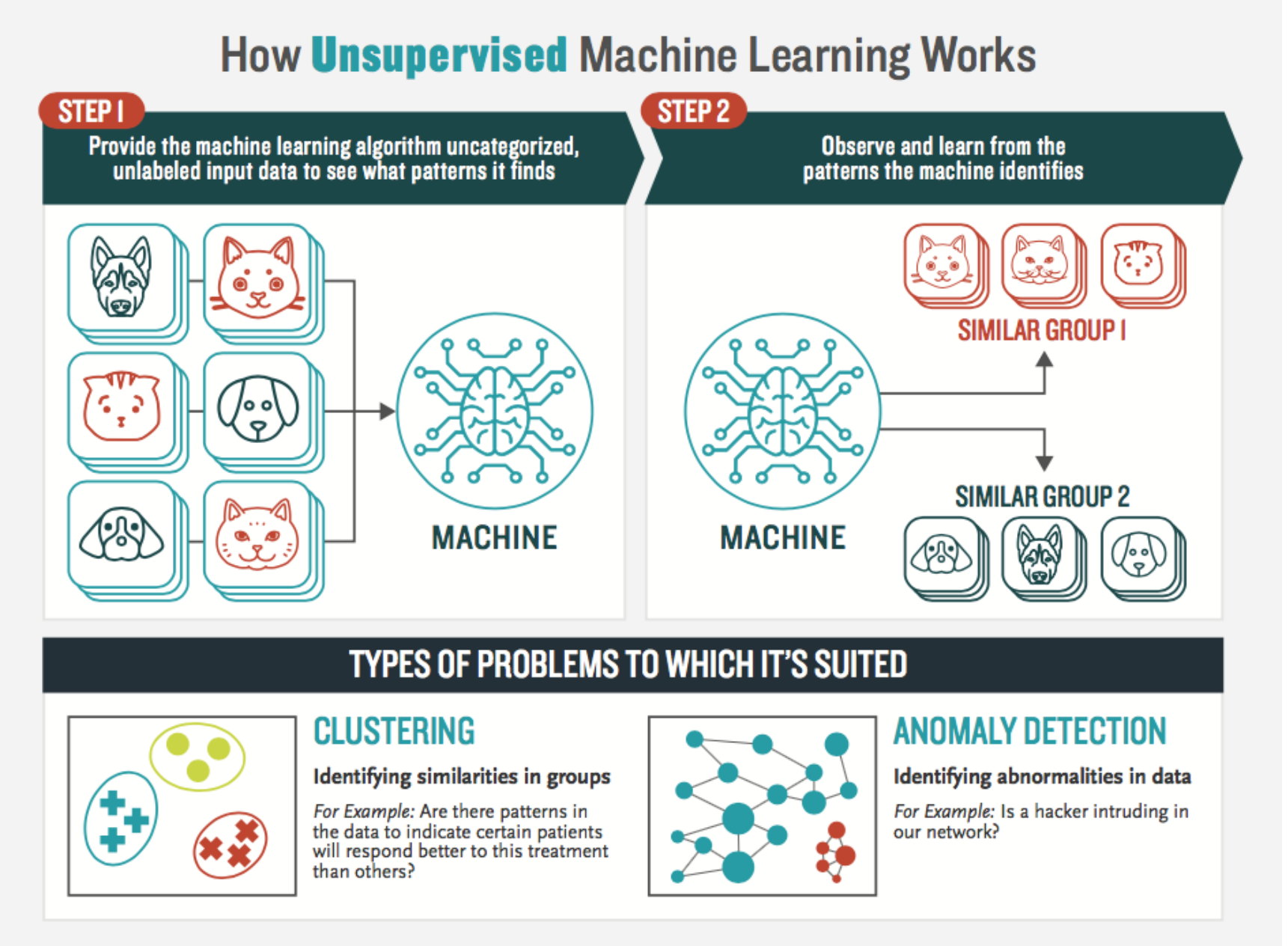
and

Unsupervised Machine Learning
Uses an unlabeled set (“That behavior looks different”)

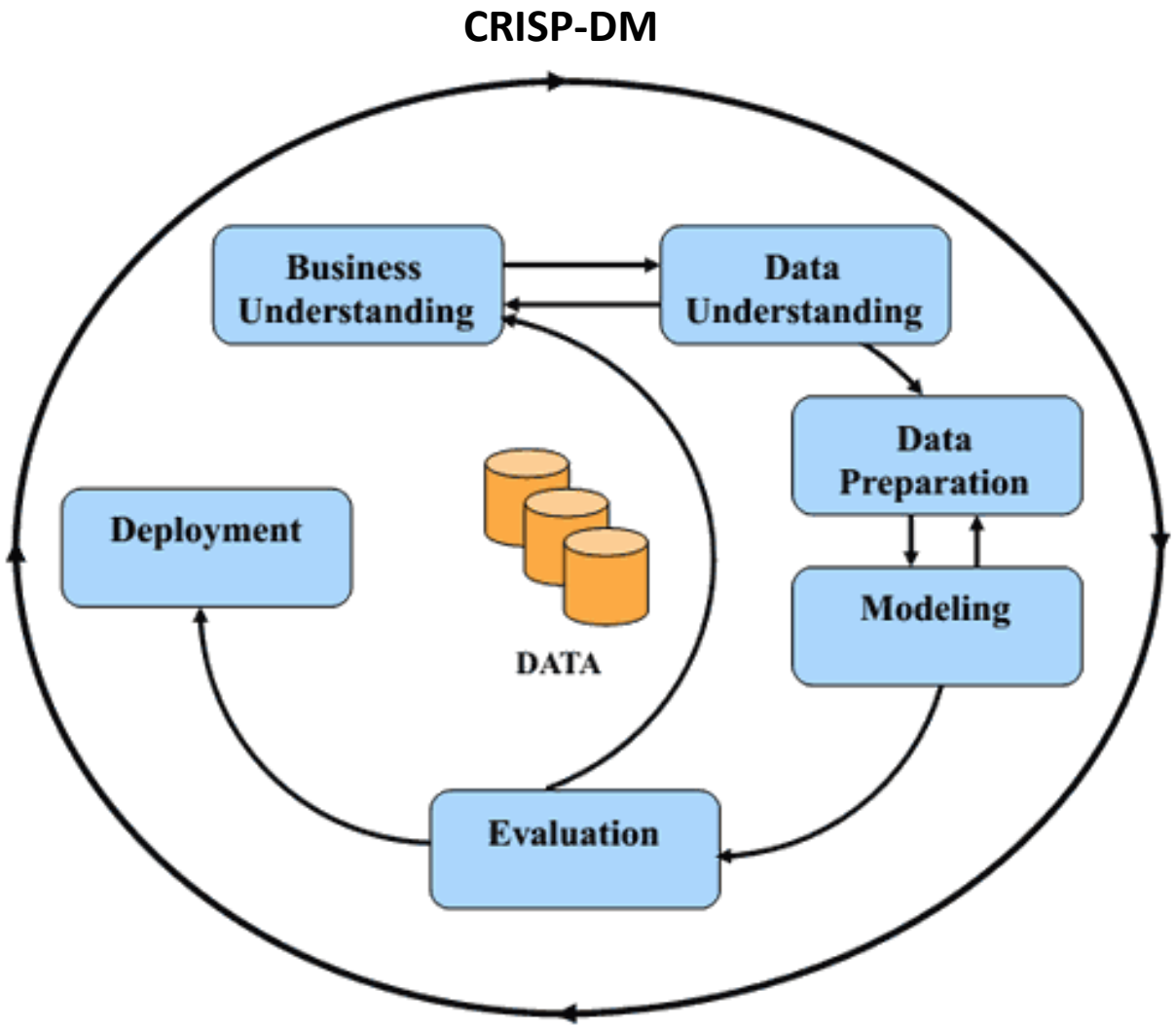
Supervised Machine Learning



Unsupervised Machine Learning

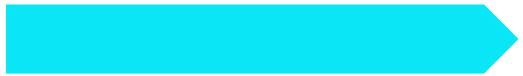
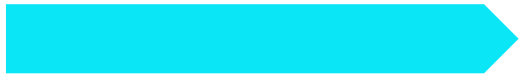
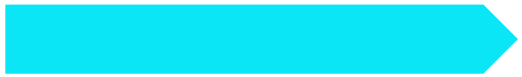


Building and Deploying Machine Learning Models Should Follow a Disciplined Approach



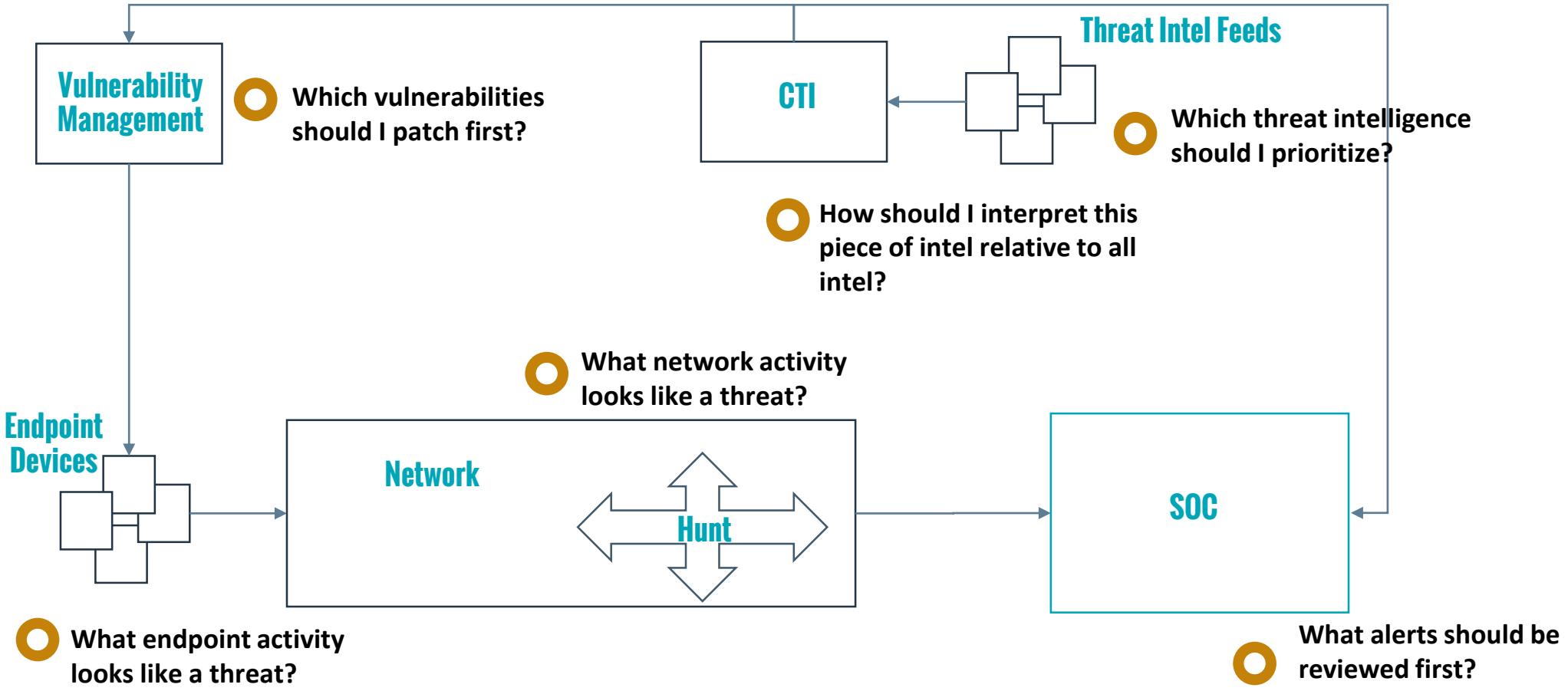
Source: MLWiki.org

SOME COMMON MACHINE LEARNING MYTHS

Myth		Truth
The model learns on its own so there is not much my people have to do		It's far from autonomous – a model will just "learn" relationships but it needs direction and data
The model tested at 99% accuracy so it will mostly be right		Very high accuracy is typically a sign of over-fitting – even the best well-fit models will make errors
Once the model has been trained and is operating I'm good to go!		A machine learning model represents a predictive relationship at the time it was trained, as conditions change your model can get worse

Moving Forward

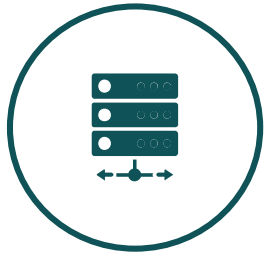
START BY IDENTIFYING YOUR MAJOR DECISION POINTS



Security decisions are disruptive and restrictive - disruptive because you're fixing something, restrictive because you're constraining behavior

Source: Network Security Through Data Analysis: Form Data to Action by Michael Collins

THE REALITIES OF APPLYING MACHINE LEARNING IN CYBER SECURITY ARE…COMPLICATED



BENIGN NETWORK ACTIVITY IS ALMOST NEVER NORMAL

Finding anomalous activity requires an understanding of what is normal, and network traffic is almost never normal

ADVERSARIES AND THEIR TACTICS ARE MOVING TARGETS

Machine learning assumes future data follow the patterns of past data, but networks and adversaries constantly change

EVERY FALSE POSITIVE COSTS TIME AND MONEY

False positives require analysts to examine an alert only to determine it was triggered by benign activity

INSIGHTS MUST BE BOTH ACCURATE AND ACTIONABLE

SOC operators need to know why a detection occurred, and black-box models can't provide that

THE RULES FOR IDENTIFYING MACHINE LEARNING APPLICATIONS IN YOUR CYBER SECURITY BUSINESS

Cyber Machine Learning Solutions Should

- ☑ Address tightly defined well-scoped problems
- ☑ Be time-sensitive, high value, and high volume
- ☑ Integrate easily with existing workflows, tools, and architecture
- ☑ Have available data to support modeling
- ☑ Have a reduced cost of false negatives / positives
- ☑ Allow for frictionless performance evaluation

WHAT TO CONSIDER WHEN DEVELOPING YOUR OWN ML PORTFOLIO

Description

- **Name**
- **Plain English Description**
- **Resources**

Technical

- **Data sources**
- **Data manipulation/EDA requirements**
- **Algorithms**
- **Written Code/Pseudocode**

Relational

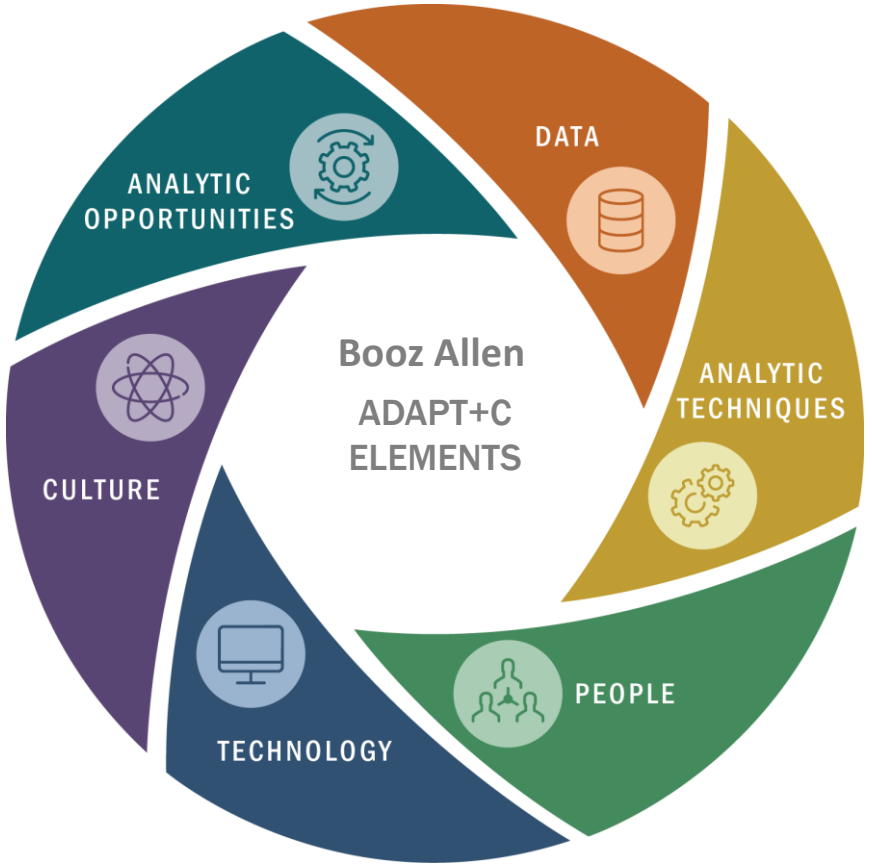
- **Associated MITRE ATT&CK Tactics**
- **Analytic Family**

Client Considerations

- **Implementation Considerations**
- **Required outputs/visualizations**
- **Key questions**
- **Etc. (Case dependent)**

AN EFFECTIVE AND SUSTAINABLE MACHINE LEARNING AND ANALYTICS CAPABILITY IS BALANCED ACROSS SIX DOMAINS

- ANALYTIC OPPORTUNITIES: Defining analytics use cases to improve organization, mission, and operations
- DATA: Using new and existing data sets to better manage and govern data
- ANALYTIC TECHNIQUES: Applying analytic tradecraft and techniques to generate insights from data
- PEOPLE: Developing talented and capable team of analytics practitioners to deliver on analytics goals
- TECHNOLOGY: Using existing and new technologies, tools, and data platforms to perform analytics projects
- CULTURE: Communicating, sharing and reinforcing the value of analytics to change staff behavior



STORIES FROM THE FIELD

Global Investment Bank

USING MACHINE LEARNING TO DETECT AND CLASSIFY BRUTE FORCE ATTACKS

Global Investment Bank

APPLYING ANOMALY DETECTION TO IDENTIFY ACCOUNT MASQUERADING

Global Auto Manufacturer

BUILDING A SECURITY DATA LAKE TO MODEL AND DETECT DATA EXFILTRATION

Global Insurance Company

USING CLUSTER ANALYSIS TO ENRICH FRAUD INVESTIGATIONS

THANK YOU!