

Watson, Neural Networks

AI'S ROLE IN NEW CYBER SECURITY FRONTIER



Sheik Sahib

CyberSecurity Architect
North American Security
ssahib@ca.ibm.com

Oct 2018



The case for AI powered CyberSecurity



Canada: Scale and frequency of cyberattacks is increasing

“The Canadian government's computer networks have been hit by state-sponsored cyberattacks about **50 times a week** — and **at least one of them usually succeeded.**”

“Between 2013 and 2015, the Government of Canada detected, on average a year, more than *2,500 state-sponsored cyber activities against its networks.*”



State-sponsored cyberattacks on Canada successful about once a week



Report says Canada not doing enough to fend off intruders, especially in vulnerable private sector

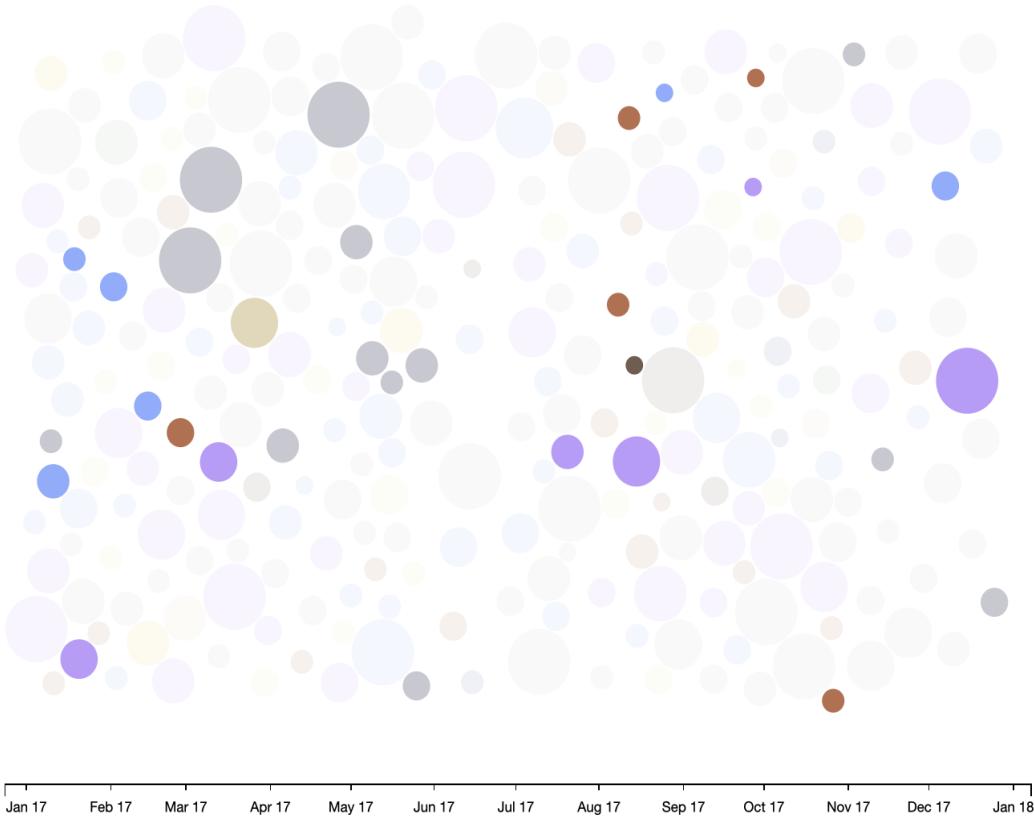


Dean Beeby · CBC News · Posted: Oct 30, 2017 5:47 PM ET | Last Updated: October 31, 2017

“Canada successfully blocks some **600 million attempts each day** to identify or exploit vulnerabilities in its government computer networks. But the vast majority are small-time hackers or other players not aligned with foreign states.”

(In 2016, 2017) .. “CSE can say that the number of cyberattacks has gone up, and that **trend is expected to continue.**”

High-profile Government Security incidents in 2017



Security Incidents < 2017 >

Displaying 279 incidents
Jan 1, 2017 to Dec 25, 2017.

Featuring 32 incidents from:
Industry: Government
([show all incidents](#))

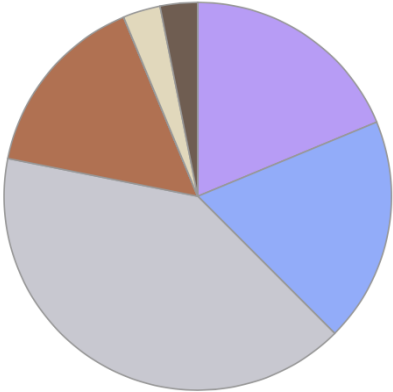
Learn More:
[About human error in security incidents](#)

Size of circle estimates relative impact of incident in terms of cost to business.

\$ \$ \$ \$\$\$\$

Attack Types (reset)

Click to view incidents for a specific attack type.

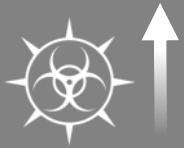


- Misconfig
- Malware
- SQLi
- Undisclosed
- DDoS
- Phishing
- Watering Hole
- Physical
- Brute Force
- Malvertising

Source: <https://www.ibm.com/security/resources/xforce/xfisi/>

Is this really sustainable?

Quick Insights: Current Security Status



Threats



Alerts



Available analysts



Needed knowledge



Available time



SKILLS
SHORTAGE

By 2022, there will be

1.8 million

unfulfilled cybersecurity jobs

Today's reality: Do all of this in <20 minutes, all day, every day

Review security incidents in SIEM

Decide which incident to focus on next

Review the data that comprise the incident (events / flows)

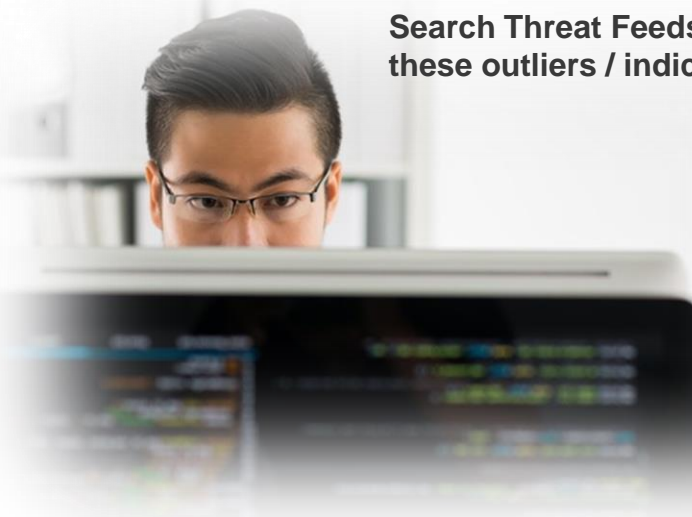
Identify the name of the malware

Pivot the data multiple ways to find outliers (such as unusual domains, IPs, file access)

Take these newly found IOCs from the internet and search from them back in SIEM

Expand your search to capture more data around that incident

Find other internal IPs are potentially infected with the same malware



Search Threat Feeds + Search Engine + Virus Total + your favorite tools for these outliers / indicators; Find new malware is at play

Start another investigation around each of these IPs

Review the payload outlying events for anything interesting (domains, MD5s, etc.)

Search more websites for IOC information for that malware from the internet

Smart but not cognitive

Go

show pictures of everything

Google Search

Meet the

Google

show pictures of everything except pink elephants

All Images Shopping Videos News More Settings Tools

About 98,800,000 results (0.62 seconds)

Images for everything except pink elephants

→ More images for everything except pink elephants

Report images

Cognitive computing enables systems to process and act on data, like humans

Understand

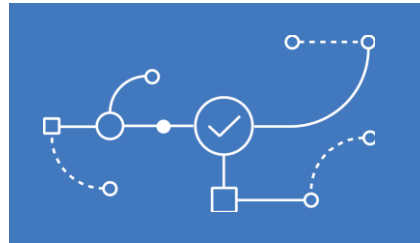


They understand

- Intent, tone, personality
- Submissions, contracts, claims
- Legal & regulatory obligations, guidelines
- News, market data...

like humans do

Reason



They can

- Identify similar risks and claims
- Assess risk
- Check for compliance
- Spot new sales opportunities, ...

infer and extract ideas

Interact



With abilities to see, talk and hear they can support

- Clients, agents & broker
- Contact center agents
- Underwriter
- Claims handler and many others

in a natural way

Learn



They learn from every interaction and

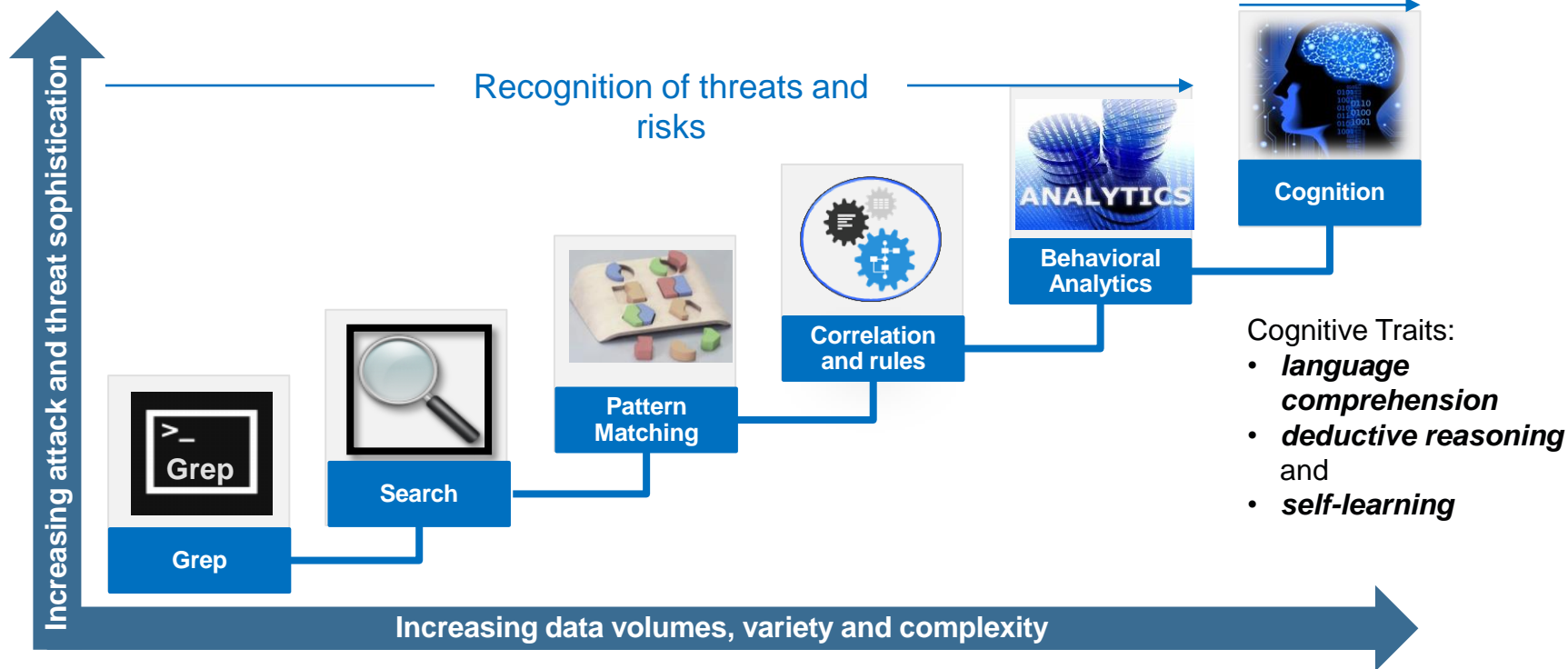
- Extract and improve best practices
- Digest new regulatory requirements, guidelines...

and never stop learning

Cognitive Solutions Reason and Present their Reasoning Process

Helping security teams not only detect where the threat is but also resolving the what, how, why, when and who to improve the overall incident response timeline

Reasoning about threats and risks

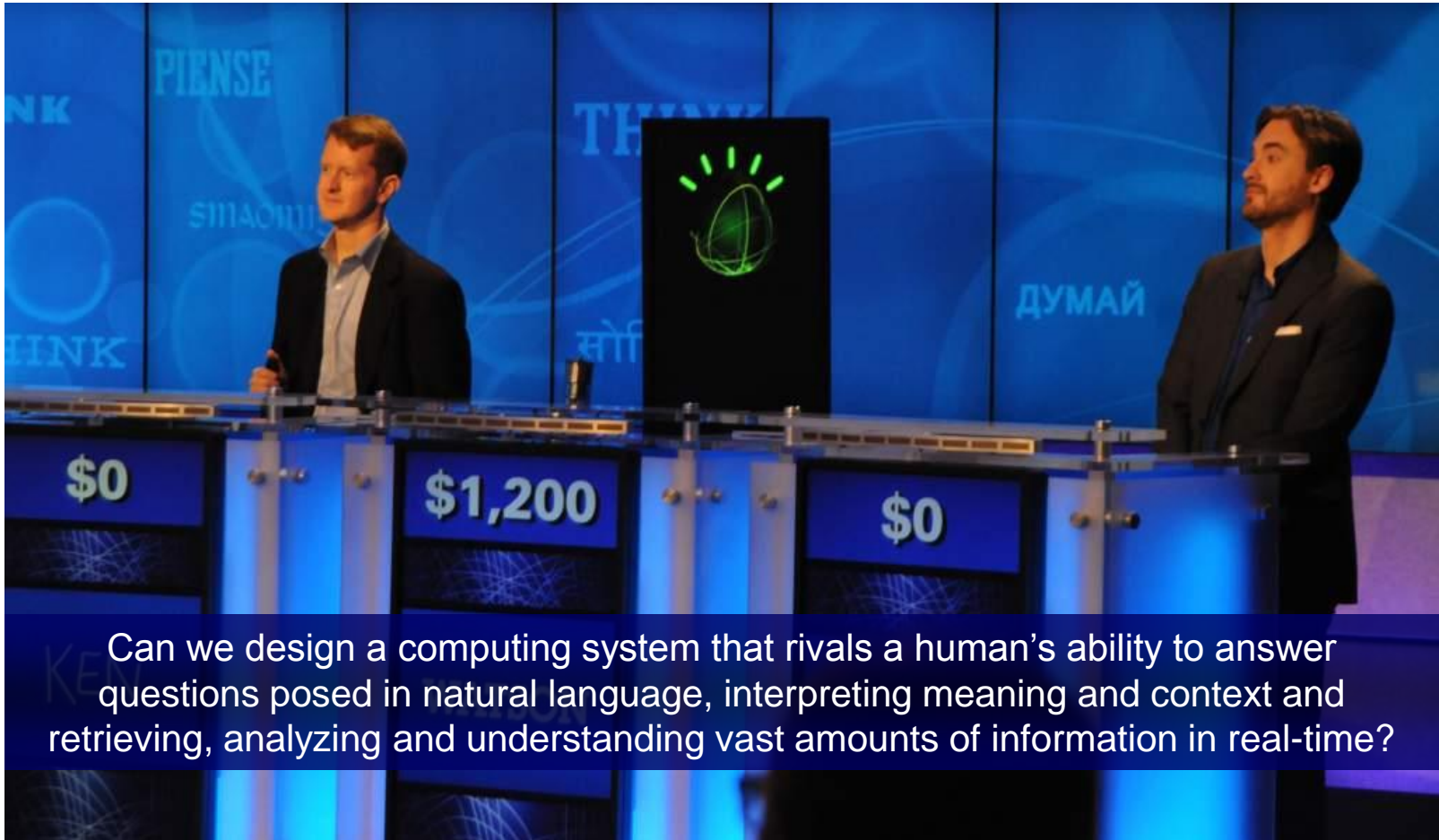




Watson AI



Watson answers a grand challenge



Can we design a computing system that rivals a human's ability to answer questions posed in natural language, interpreting meaning and context and retrieving, analyzing and understanding vast amounts of information in real-time?

A Computer Called Watson



A COMPUTER CALLED WATSON



Watson is an efficient analytical engine that pulls many sources of data together in real-time, discovers an insight, and deciphers a degree of confidence.

IBM Watson on IBM.com

In an historic event, in February 2011 IBM's Watson computer competed on *Jeopardy!* against the TV quiz show's two biggest all-time champions. Watson is a computer running software called Deep QA, developed by IBM Research. While the grand challenge driving the project was to win on *Jeopardy!*, the broader goal of Watson was to create a new generation of technology that can find answers in unstructured data more effectively than standard search technology.

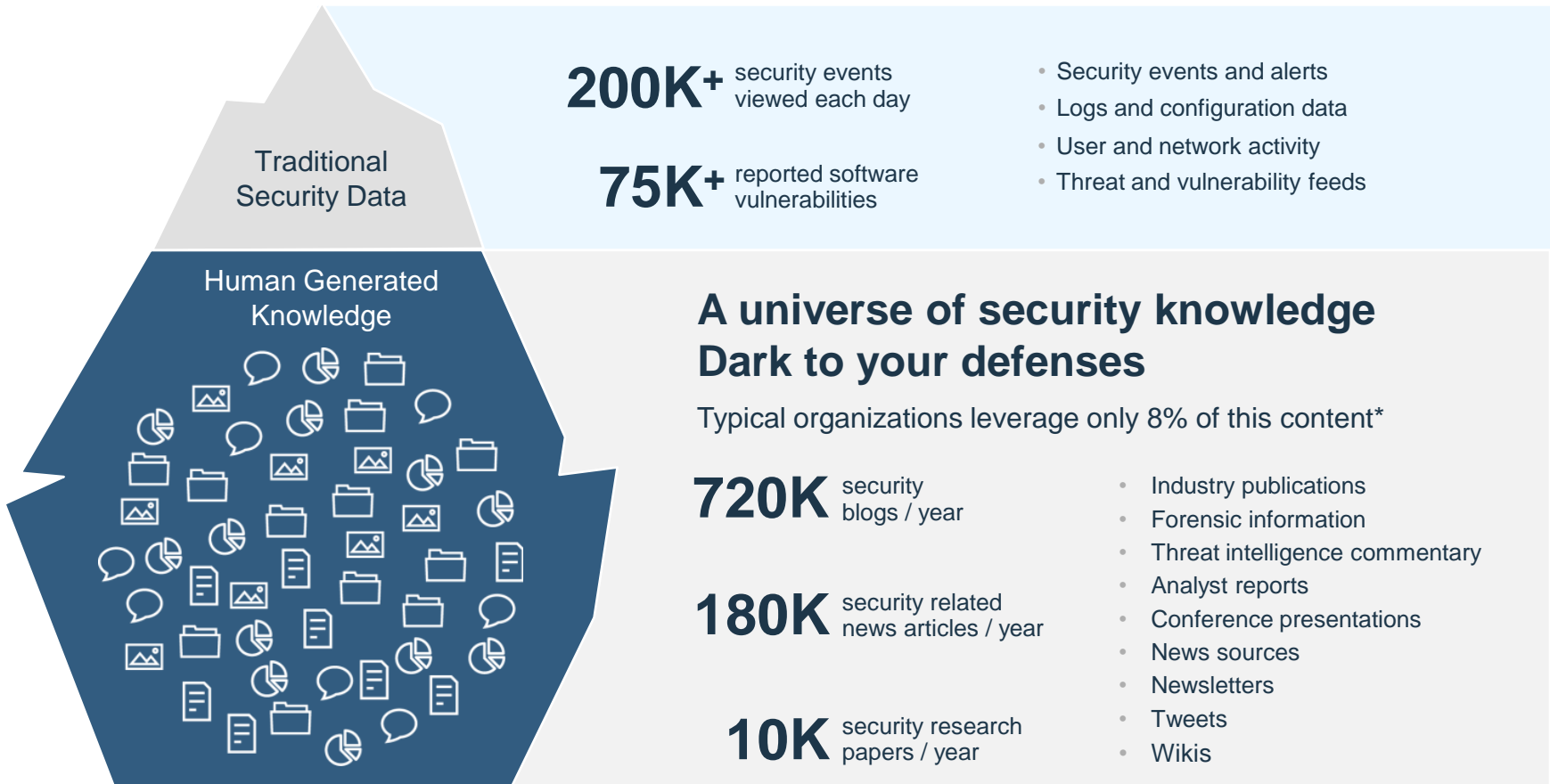
Final Score: Rutter - \$21,600 Jennings - \$24,000 **Watson - \$77,147**



Cybersecurity powered by AI

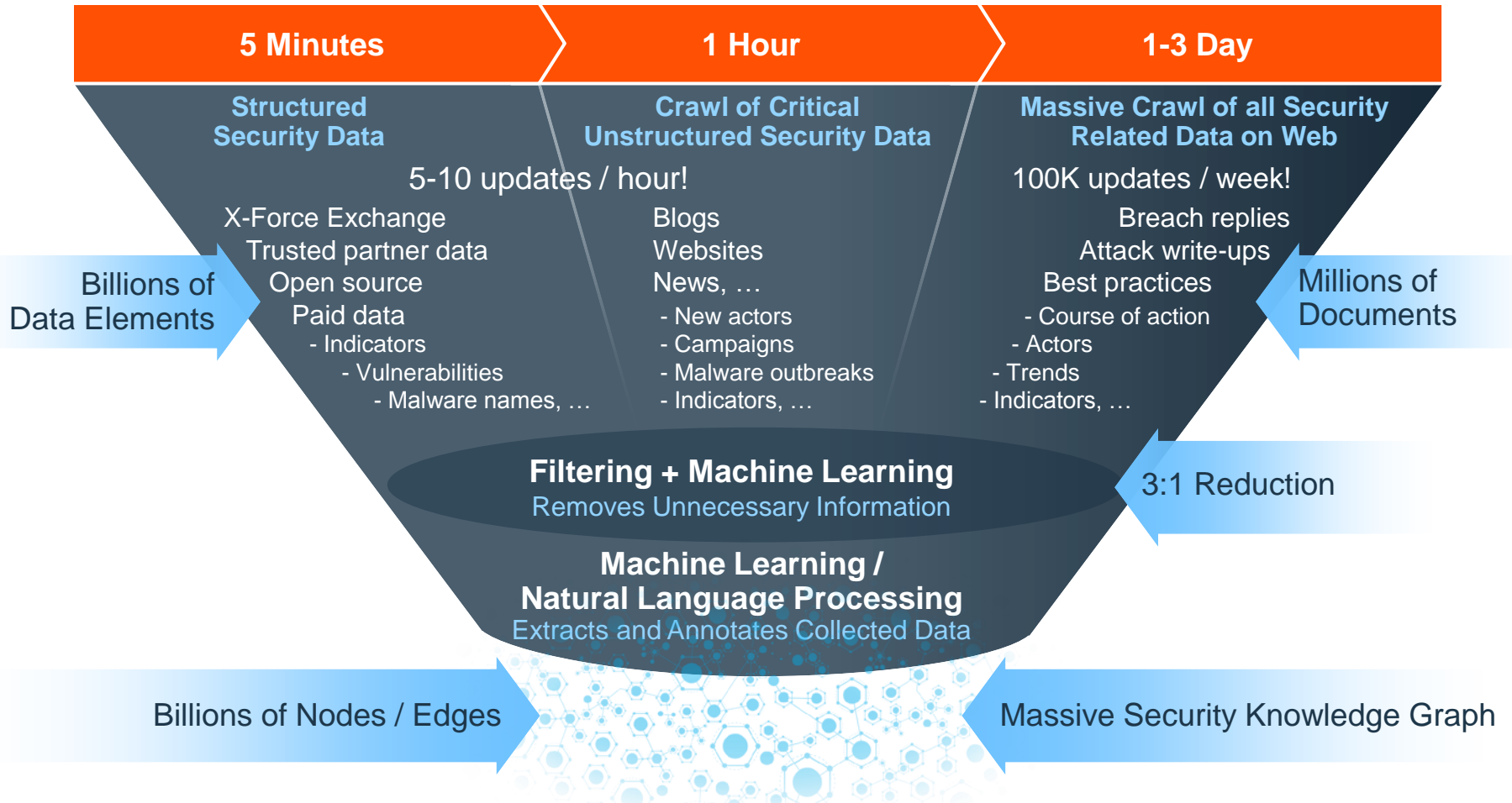


A tremendous amount of security knowledge is created for human consumption, but most of it is untapped



¹ Forrester Research : Can You Give The Business The Data That It Needs? , 2013

Cognitive Security unlocks vast security knowledge to quickly enable comprehensive investigative insights



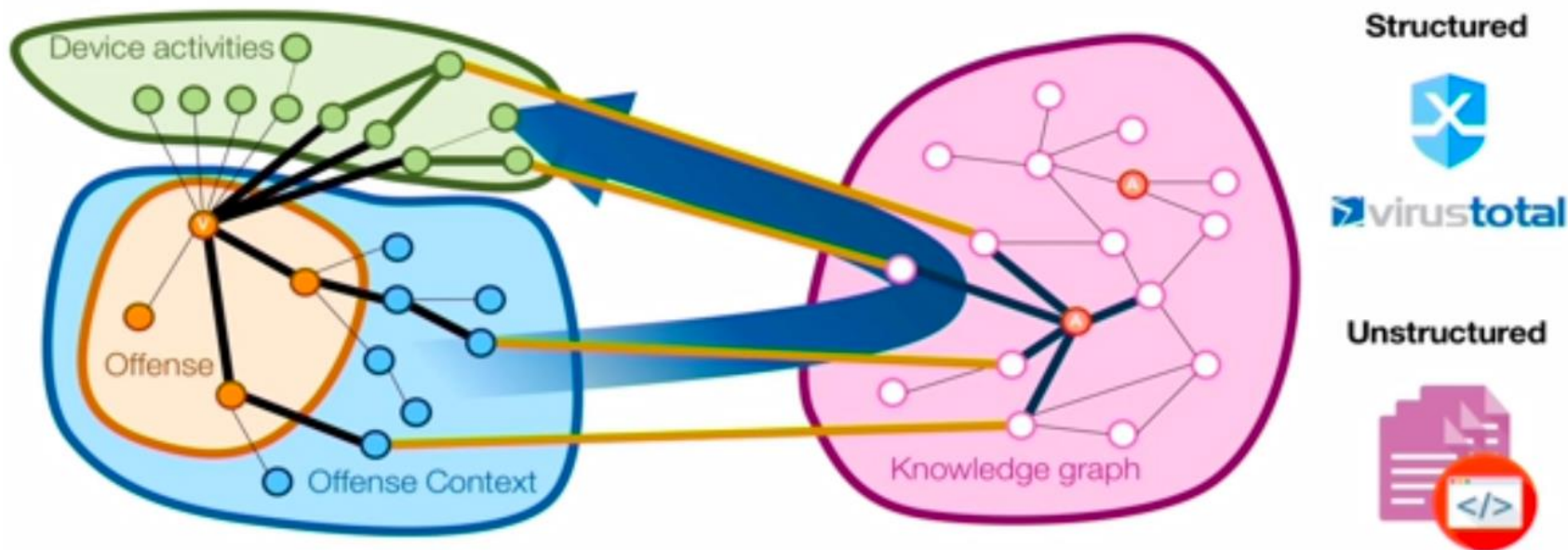
Connecting the dots

Offense Context Graph

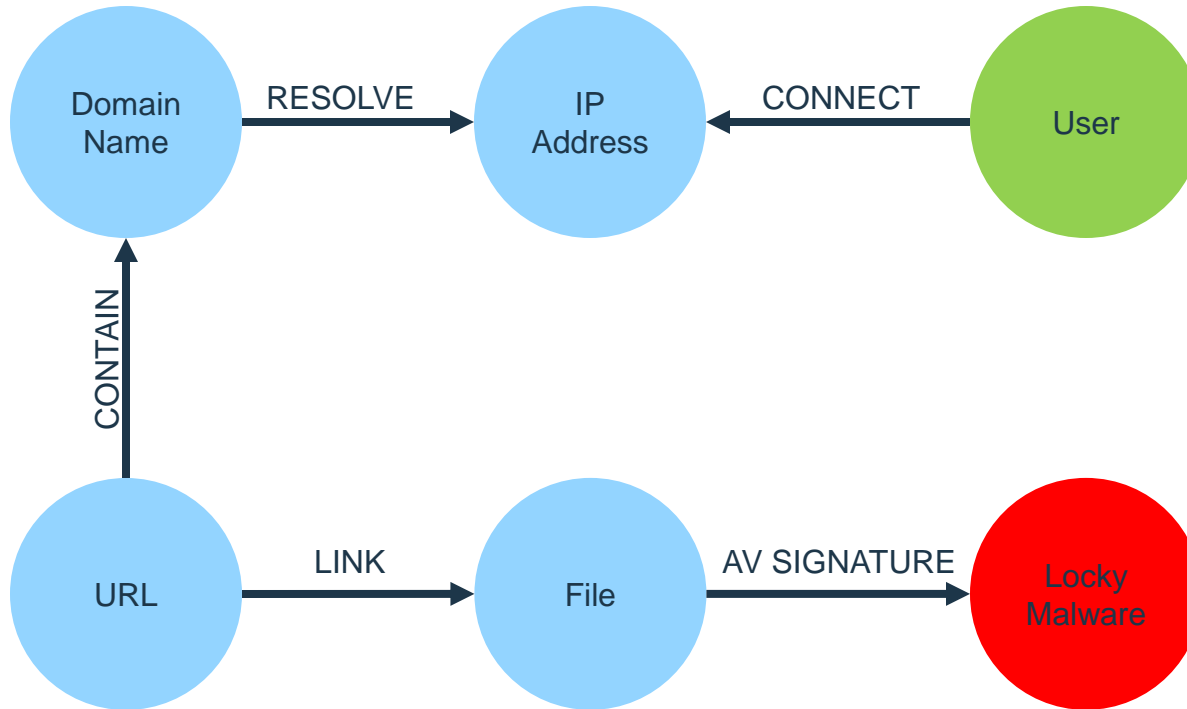
Local Kinetics and Event Data

Security Knowledge Graph

Security and Threat Information



Connecting the dots – an example



Cognitive systems bridge this gap and unlock a new partnership between security analysts and their technology

Human Expertise

- Common sense
- Abstraction
- Morals
- Dilemmas
- Compassion
- Generalization

Security Analytics

- Data correlation
- Pattern identification
- Anomaly detection
- Prioritization
- Data visualization
- Workflow



Cognitive Security

- Unstructured analysis
- Natural language
- Question and answer
- Machine learning
- Bias elimination
- Tradeoff analytics



AI Cybersecurity in the real world .. *IBM perspective*



Using Artificial Intelligence to address growing security needs

Predictive Analytics

- **Approach:** Model behaviors and identify emerging and past threats and risks
- **Applications:** Network, user, endpoint, app and data, cloud



A IBM QRadar User Behavior Analytics

Intelligence Consolidation

- **Approach:** Curation of intelligence and contextual reasoning
- **Applications:** Structured and unstructured (NLP) data sources



B IBM QRadar Advisor / Watson for Cybersecurity

Trusted Advisors & Response

- **Approach:** Reason about security events for triage and response
- **Applications:** Cognitive SOC analyst, orchestration, automation and digital guardian



C Take action with QRadar User Behavior Analytics

Predictive analytics across IBM Security portfolio

What we predict...	Product	Models used	Inputs	Output
Insider Threats	QRadar UBA	Peer grouping, time-series, anomaly	Security logs and events	Risk score of users
Malicious Traffic	QRadar Network Insights	Random forest	Network data	Risk score of flows
Botnet Domains	X-Force DNS Analytics QRadar DNS Analytics	Multiple	DNS data, registrar info	Domain risk score and reputation
Vulnerable Code	AppScan Intelligent Code / Findings	Random forest, logistic regression	Scans from benchmark set of applications	New vulnerability rules, reduced false positives
Database Attacks	Guardium Outlier Detection	Anomaly, user and DB cluster	Sql queries, errors, file access activity	Abnormal activity, hourly risk score
Risky User Access	IAM Governance, Authentication	Outlier detection with peer group	IAM data, logs and UBA alerts	Risk score of users, apps
Fraudulent Users	Trusteer Behavioral Biometrics	Random forest	Keystrokes, app, mouse usage	Risk score of users
Phishing Websites	Trusteer Cognitive Phishing	Random forest	URLs and website content	Risk score of suspected sites

Intelligence consolidation and Trusted Advisors

What we do...	Product	Models used	Inputs	Output
Security intelligence consolidation	Watson for Cybersecurity	Watson Natural Language Understanding	Unstructured content, web content	Cybersecurity contextual knowledge base
Automatic offense investigations	QRadar Advisor	Multiple	QRadar events	Root cause analysis, augmented context
Virtual Cybersecurity Analyst	IBM Havyn	Watson Speech	Voice, unstructured content, threat content	Contextual security information, spoken content
Mobile endpoint management advisor	MaaS360 Advisor	Watson	Unstructured content, threat alerts, etc.	Personalized mobile endpoint management recommendations
Mobile end-user self-service assistant	MaaS360 AI Assistant	Watson Speech	User commands, calendar and email contents, support knowledge base	Coordinates calendar and email activities; provides real-time end-user support

Cognitive: Revolutionizing how security analysts work

- Natural language processing with security that understands, reasons, learns, and interacts

The screenshot displays the IBM QRadar Security Intelligence interface. On the left, the 'Analyzer' sidebar shows 'Offense 1' with details: Type: Source IP, Last Update: Today at 12:00 PM, Assigned to: Magnitude: 5. Below this is a list of 'Observables' with counts: AV Signature (2), Domain (2), Endpoint (2), File (21), Filename (2), Hash (20), IP (17), Malware (1), Reputation (30), and URL (20). The main area shows a network graph with a central IP node (18.103.122.188) connected to various other nodes including malware signatures like 'Trojan, Win32, ...' and 'Wextract'. A detailed incident report window is overlaid on the right, titled 'Incident 1: Potentially Successful Exploit containing TCP_HIT'. It shows a score of 5 and a 'Watson' analysis. The report includes a table of indicators: Malware Family (1), AV Signature (2), Filename (23), and Url (22). Below this is a 'Watson Insights' section with a quote: 'QRadar Advisor's analysis of 117 observables from this offense has finished. The reasoning process discovered 283 new indicators that were not part of the offense. A total of 110 data points have been found to be linked with the offense. 34 of all indicators are known to be related with suspicious activity, two of them have been observed actively in this offense. From the newly found indicators, 32 have ties to suspicious activity. In particular, 19 URLs and 15 IP addresses have been found, which are known to be suspicious or malicious. The following malware family type may be linked to the offense: smokeloader.' The report also includes 'Supporting Details' with a bar chart for Reputation (25%) and Filename (19%), and a donut chart.

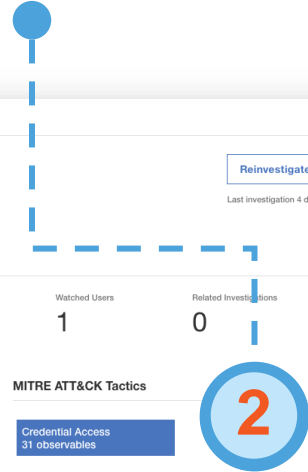
Watson determines the specific campaign (Locky), discovers more infected endpoints, and sends results to the incident response team

Cognitive: Aligning incidents to the ATT&CK chain

Confidence level for each progression validates the threat



Visualize how the attack has occurred and is progressing



Uncover what tactics can still possibly occur



Watson Investigations / ID: Offense 126

Key findings for **Source IP 192.168.0.119**
Default | Investigated

Reinvestigate | Graph Relationships
Last investigation 4 days ago, on October 4, 2018, 8:50 PM

Concern Medium

Key Insights

Threat Actors	Malware Families	High Value Assets	Risky Users	Watched Users	Related Investigations	Duplicate Investigations
0	5	2	0	1	0	0

Key Observables

Total	Suspicious	Critical	New Local Context
65	44	35	4

MITRE ATT&CK Tactics

Credential Access
31 observables

Insights

From this offense, Watson has analyzed 24 observables. The analysis found 73 new indicators that were not included in the offense. A total of 35 data points were found to be linked with the offense. 31 indicators were related to suspicious activity, and six indicators were active. From the newly found indicators, 25 have ties to suspicious activity. In particular, four files, 24 URLs, one domain name and two IP addresses have been found, which are known to be suspicious or malicious. The following malware family types might be linked to the offense: "icepack", "locky", "dridex", "spam zero-day", "emotet". The evidence is provided by "three anti-virus signatures". One user on a watch list is associated with the offense: kyle.langford. Advisor has identified one high value asset associated with the offense: 192.168.0.119.

Analysis of the indicators found by Watson revealed four additional observables related to the offense in the local context. Advisor has identified one additional high value asset related to the offense in the expanded local context: 192.168.0.122.

Offense Summary | View details

IBM QRadar UBA: Machine Learning Algorithms

IBM QRadar Security Intelligence | admin | Help | Messages 11 | IBM | System Time: 8:05 PM

Dashboard > User Details | Search for User | Reset Layout

Grant Lamson

Grant Lamson
Grant_Lamson@example.ibm.com
491865
IT Support Specialist
Support
Yokohama, Japan

Risk Events Generated by User (Last Hour)
5
Overall Score: 531 ↓
Current Score: 5

Risk Score

840
600
400
200
May 1 12:00 May 2 12:00 19:00

Sense Events

May 2, 2017 +715
5 UBA-ML : Abnormal increase in User activity +500

User Activity by Category

May 2, 6:00 PM

Access
System
Suspicious App
Sense
Application
Authentication

Actual (Grey)
Learned (Blue)

'Sense' Activity

May 1 - May 2

15
10
5
0
13:00 14:00 15:00 16:00 17:00 18:00

Actual (Grey)
Learned (Blue)

Recent Offenses

Offense # 5283 a day ago
Grant Lamson
Categories: Access Permitted, Information, User Risk, Sense
Offense
Event Count: 14 Flow Count: 0 Magnitude: 3/10

“Deviations from normal behavior”

Adversarial AI



Attacker's Use of AI Today

ME: Model Extraction
DE: Data Extraction
Ev: Model Evasion
Po: Model Poisoning

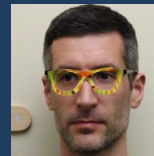
AI Powered Attacks

- **Generate:** DeepHack tool learned SQL injection [DEFCON'17]
- **Automate:** generate targeted phishing attacks on Twitter [Zerofox Blackhat'16]
- **Refine:** Neural network powered password crackers
- **Evade:** Generative adversarial networks learn novel steganographic channels



Attacking AI

- **Poison:** Microsoft Tay chatbot poisoning via Twitter (and Watson "poisoning" from Urban Dictionary) [Po]
- **Evade:** Real-world attacks on computer vision for facial recognition biometrics [CCS'16] and autonomous vehicles [OpenAI] [Ev]
- **Harden:** Genetic algorithms and reinforcement learning (OpenAI Gym) to evade malware detectors [Blackhat/DEFCON'17] [Ev]



Theft of AI

- **Theft:** Stealing machine learning models via public APIs [USENIX'16] [DE]
- **Transferability:** Practical black-box attacks learn surrogate models for transfer attacks [ASIACCS'17] [ME, Ev]
- **Privacy:** Model inversion attacks steal training data [CCS'15] [DE]



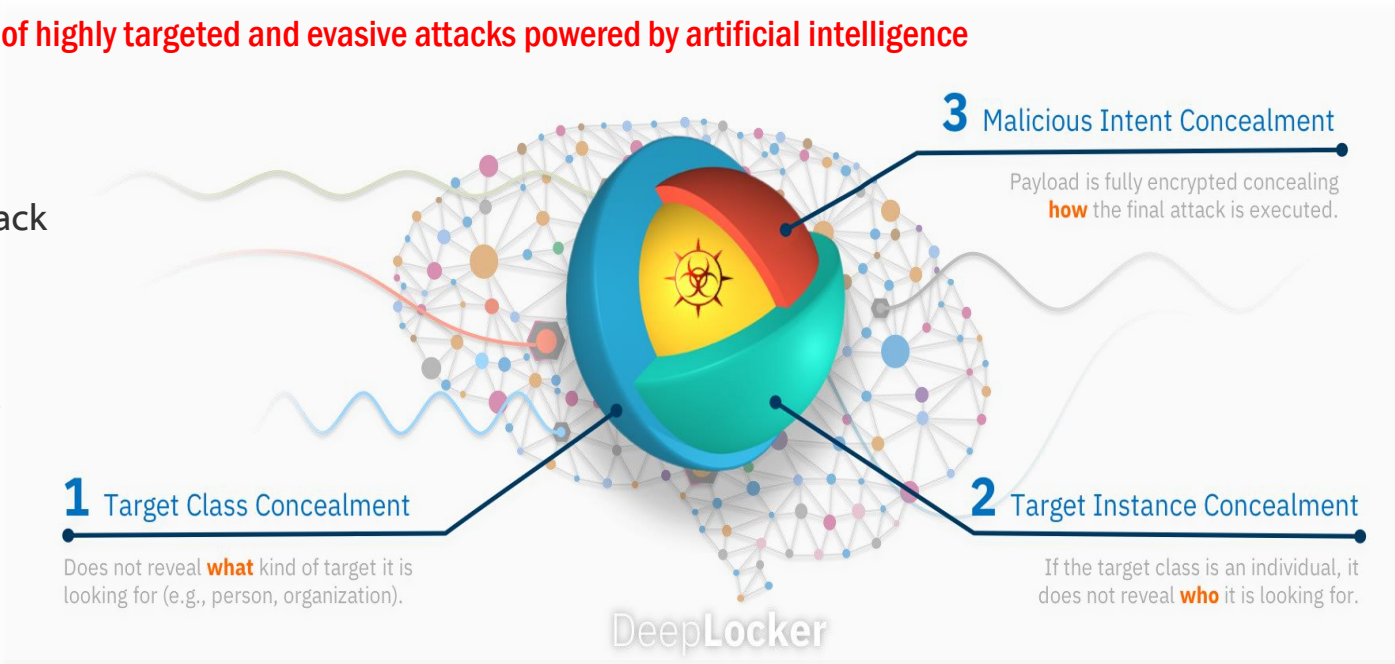
IBM Deep Locker: Concealing Targeted Attacks with AI Locksmithing

<https://www.blackhat.com/us-18/briefings.html#deeplocker-concealing-targeted-attacks-with-ai-locksmithing>

DeepLocker - a novel class of highly targeted and evasive attacks powered by artificial intelligence (AI)

A stealthy, targeted attack needs to conceal two main components:

- *trigger condition(s)*
- *the attack payload.*






- DeepLocker leverage the “black-box” nature of the *DNN AI model* to conceal the trigger condition.
- A simple “if this, then that” trigger condition is transformed into a deep convolutional network of the AI model that is very hard to decipher.
- In addition to that, it is able to convert the concealed trigger condition itself into a “password” or “key” that is required to unlock the attack payload.



THANK YOU

FOLLOW US ON:

-  ibm.com/security
-  securityintelligence.com
-  xforce.ibmcloud.com
-  [@ibmsecurity](https://twitter.com/ibmsecurity)
-  youtube/user/ibmsecuritysolutions

© Copyright IBM Corporation 2017. All rights reserved. The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. Any statement of direction represents IBM's current intent, is subject to change or withdrawal, and represent only goals and objectives. IBM, the IBM logo, and other IBM products and services are trademarks of the International Business Machines Corporation, in the United States, other countries or both. Other company, product, or service names may be trademarks or service marks of others.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

Introduction to Machine Learning

A subfield of computer science that enables computers to learn without being explicitly programmed

- Arthur Samuel in 1959

Supervised Learning

Inferring a general rule or mathematical function from labeled training data to be applied to other data

Primary Use Cases


- Regression Analysis
 - Deriving correlation relationships between variables and estimating the strength of those relationships
 - Widely used for prediction and forecasting
- Classification:
 - Produces a model from a training set that can assign unseen inputs into different categories

Unsupervised Learning

Detecting the presence of patterns or models from unlabeled data

Primary Use Cases

- Clustering
 - Data is divided into different groups based on one or more attributes
- Dimensionality Reduction
 - process of reducing the number of random variables under consideration, via obtaining a set of principal variables
 - Feature Selection: finding subset of the original variables
 - Feature Extraction: transform high-dimensional space to a space of fewer dimensions



There is a massive amount of noise out there; the human brain can't process everything on a day-to-day basis. We need something to help, something like AI or cognitive technologies.

Chad Holmes – Principal and Cyber-Strategy, Technology and Growth Leader (CTO) at Ernst & Young LLP

“Cognitive security has so much potential — you can meet your labor shortage gap, you can reduce your risk profile, you can increase your efficiency of response. It can help you understand the narrative story. People consume stories — this happened, then this happened, with this impact, by this person.

Additionally, cognitive can lower the skills it takes to get involved in cybersecurity. It allows you to bring in new perspectives from non-IT backgrounds into cracking the problem.”

David Shipley – Director of Strategic Initiatives, Information Technology Services, University of New Brunswick

Artificial Intelligence and Sub Categories

Artificial Intelligence

Cognitive

Machine Learning

Deep Learning

- Machine learning is a subfield of AI and computer science that has its roots in statistics and mathematical optimization. Machine learning covers techniques in supervised and unsupervised learning for applications in prediction, analytics, and data mining.*
- Deep learning isn't an algorithm, per se, but rather a family of algorithms that implement deep networks with unsupervised learning.*

* "A beginner's guide to artificial intelligence, machine learning, and cognitive computing"
<https://www.ibm.com/developerworks/library/cc-beginner-guide-machine-learning-ai-cognitive/index.html>

Adversarial Robustness Toolbox (ART)

IBM Research announced:

ART – an open-source library for adversarial machine learning

- ART provides an implementation for many state-of-the-art methods for attacking and defending classifiers
- ART allows rapid crafting & analysis of attacks and defense methods for machine learning models

<https://github.com/IBM/adversarial-robustness-toolbox>