Artificial Intelligence

Applications in Cyber Defense

Isabelle Andrade
https://linkedin.com/in/isabelledeandrade/

2021

Objectives

- To get to know the main concepts of Artificial Intelligence and its applications in the cyber domain
- To be able to apply these concepts in a simulated scenario

Overview

- Artificial Intelligence Concepts
- AI Cyber Defense Applications
- Hands-on Exercise

Overview

- Artificial Intelligence Concepts
- AI Cyber Defense Applications
- Hands-on Exercise

"The time was one minute past midnight. But he was the only one who had to sit on his way back. The time was one minute after midnight and the wind was still standing on the counter and the little patch of straw was still still and the street was open."

?

“It was mountainous. Route 6 came over the river, wound around a traffic circle, and disappeared into the wilderness. Not only was there no traffic but the rain came down in buckets and I had no shelter.”

"The time was one minute past midnight. But he was the only one who had to sit on his way back. The time was one minute after midnight and the wind was still standing on the counter and the little patch of straw was still still and the street was open."

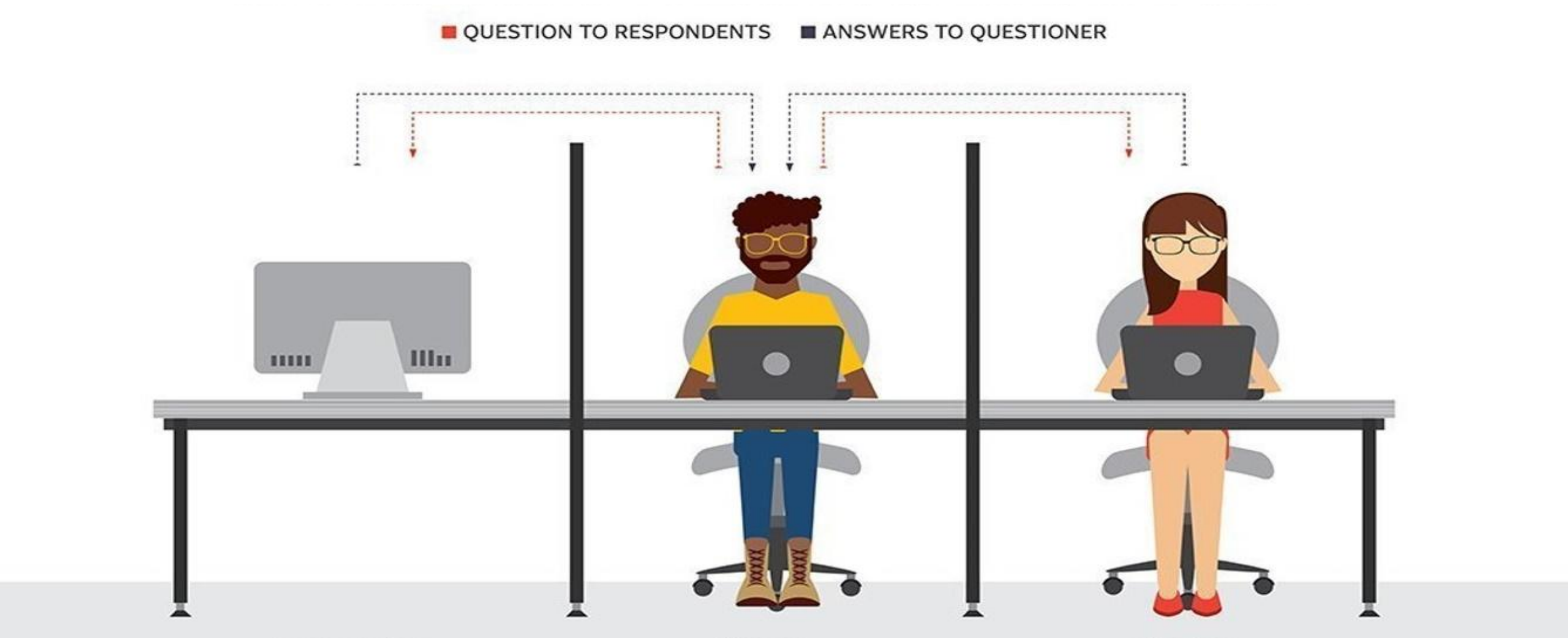
1 the Road - AI (Kenric McDowell and Ross Goodwin)

!

"It was mountainous. Route 6 came over the river, wound around a traffic circle, and disappeared into the wilderness. Not only was there no traffic but the rain came down in buckets and I had no shelter."

On the Road - Jack Kerouac

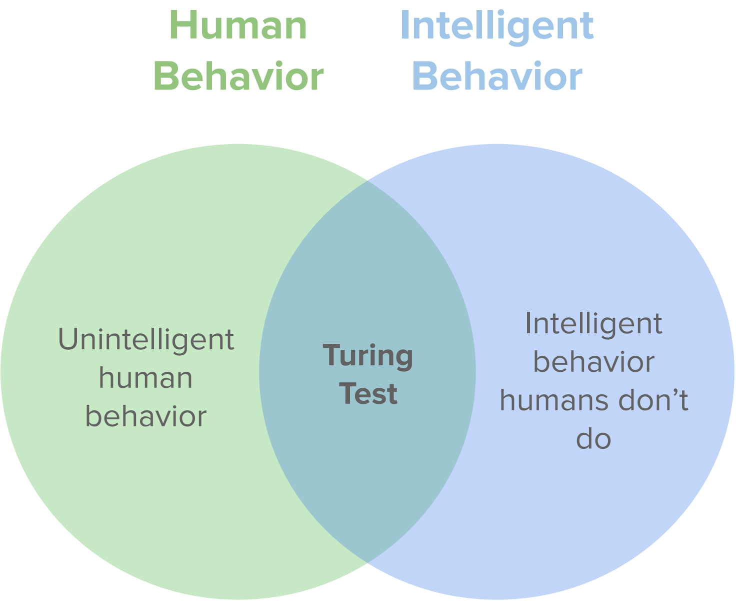
Turing Test



During the Turing test, the **human questioner** asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the **human respondent** and which terminal is operated by the **computer**

“Can machines think?”

Alan Turing, Computing Machinery and Intelligence, 1950.


Human Behavior
Intelligent Behavior
Unintelligent human behavior
Turing Test
Intelligent behavior humans don’t do

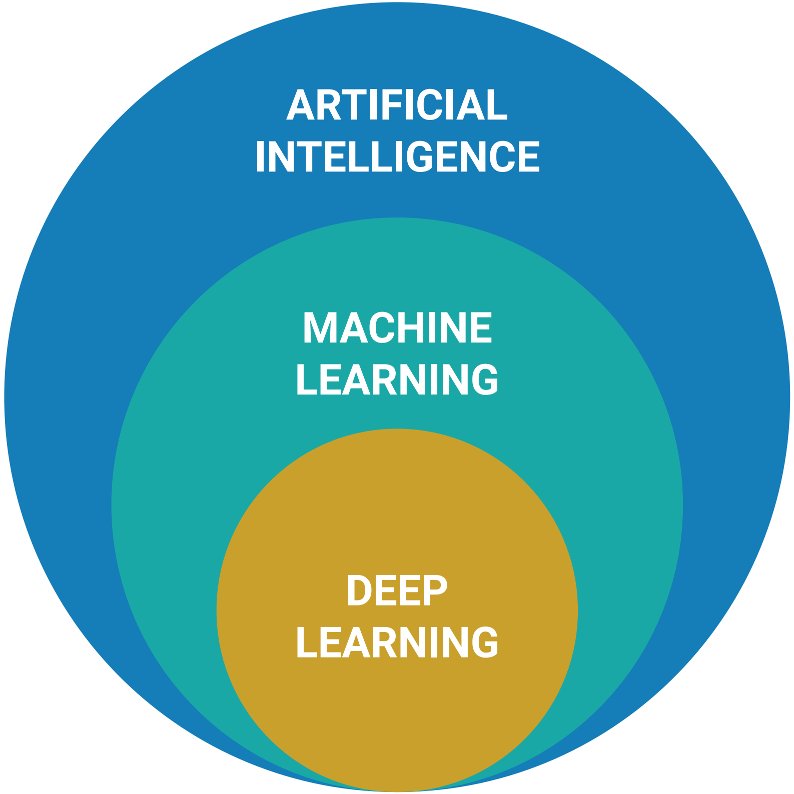
Artificial Intelligence

“The science and engineering of making **intelligent machines**.”

- John MacCarthy, Stanford University Professor, Dartmouth Summer Research Project on Artificial Intelligence, 1956

“At its core, AI is the branch of computer science that aims to **answer Turing's question in the affirmative**. It is the endeavor to replicate or simulate human intelligence in machines.”

- Builtin Staff Writers, Tech Hiring Company, 2021



Artificial Intelligence

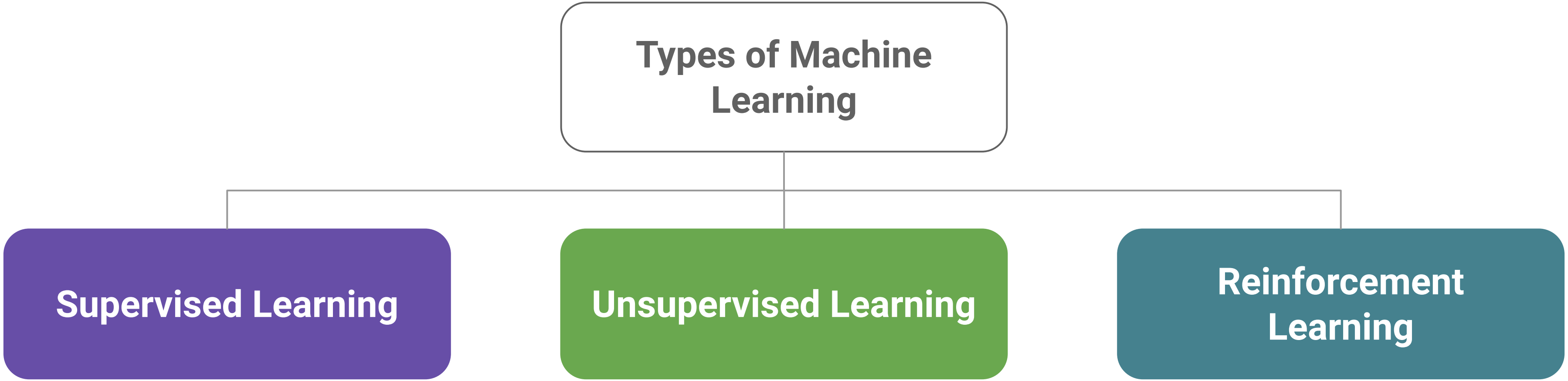
Programs with the ability to learn and reason like humans

Machine Learning

Subset of AI that uses statistical methods to enable machines to improve through experience, without being specifically explicitly programmed

Deep Learning

Subset of ML in which artificial neural networks adapt and learn from vast amounts of data



Types of Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

Task Driven

Predict new values based on labeled data

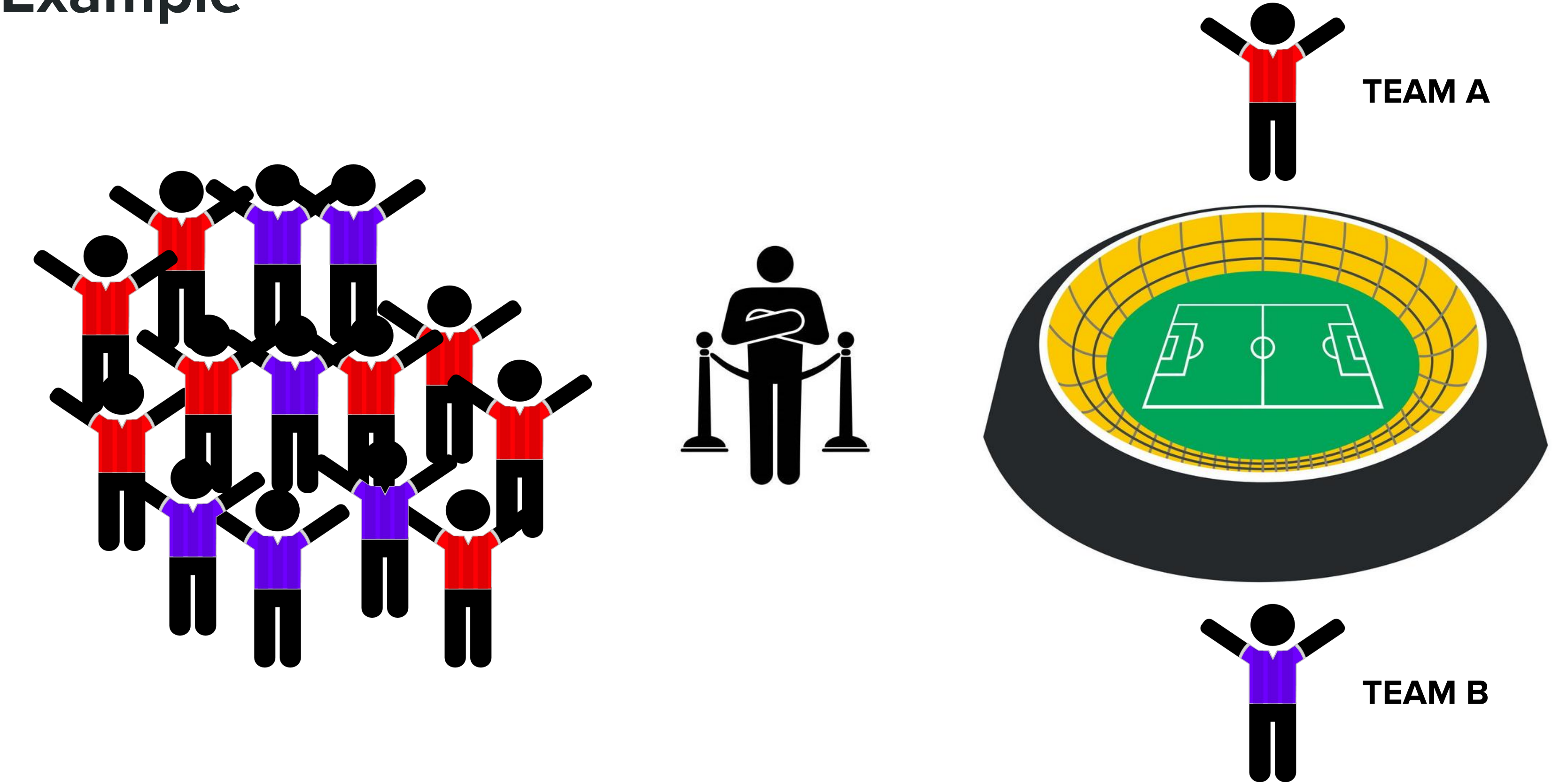
Data Driven

Find hidden structure in data, identify clusters

Reward Driven

Learn from mistakes (rewards and penalties)

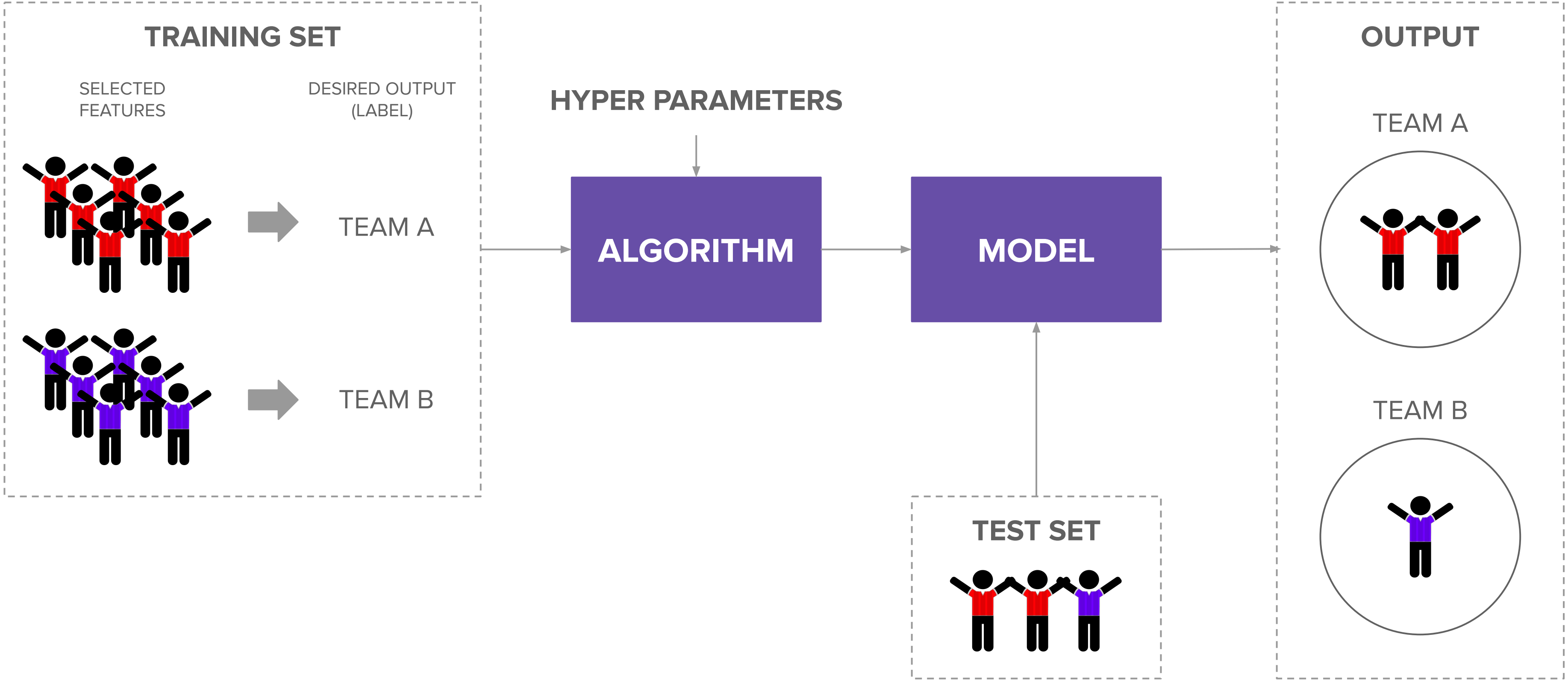
Example



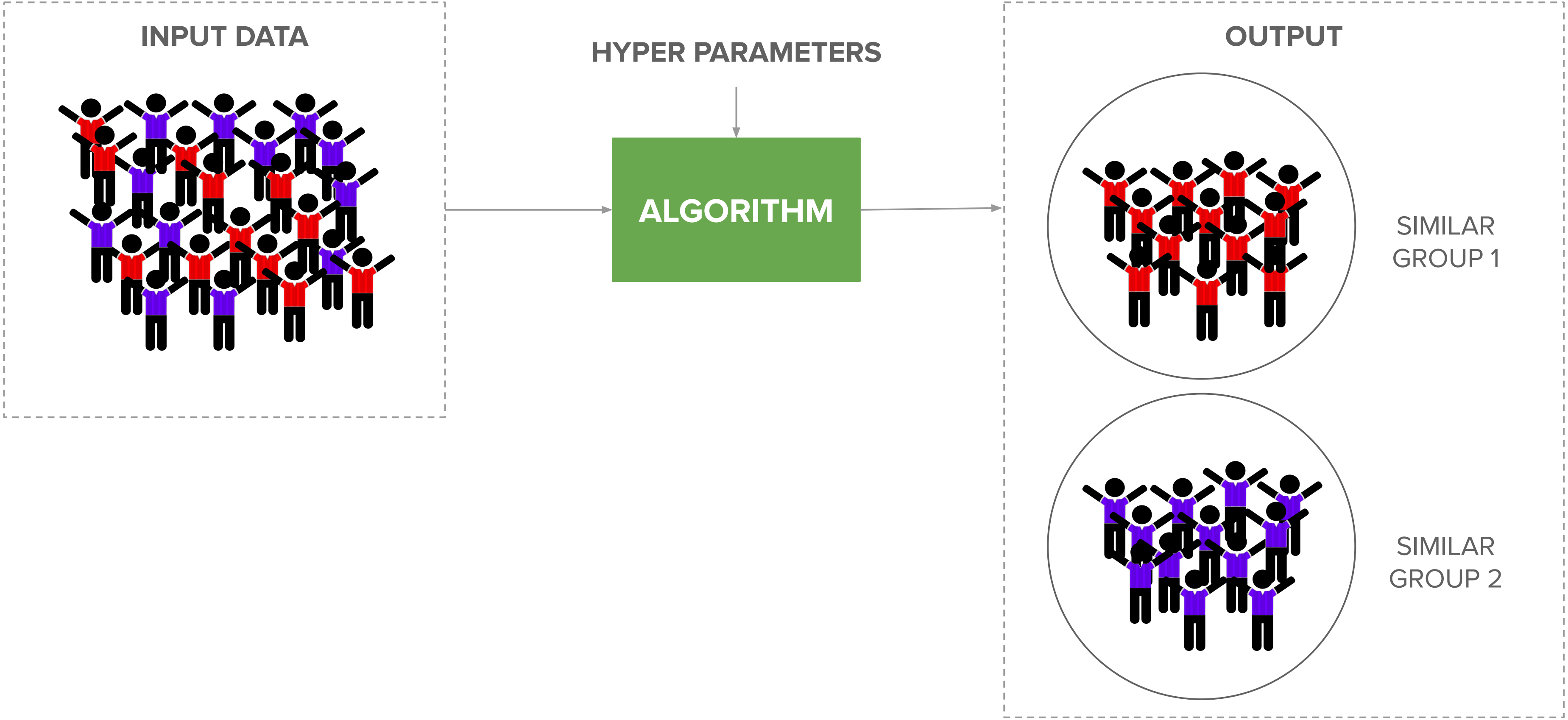
TEAM A

TEAM B

Supervised Machine Learning



Unsupervised Machine Learning



INPUT DATA

HYPER PARAMETERS

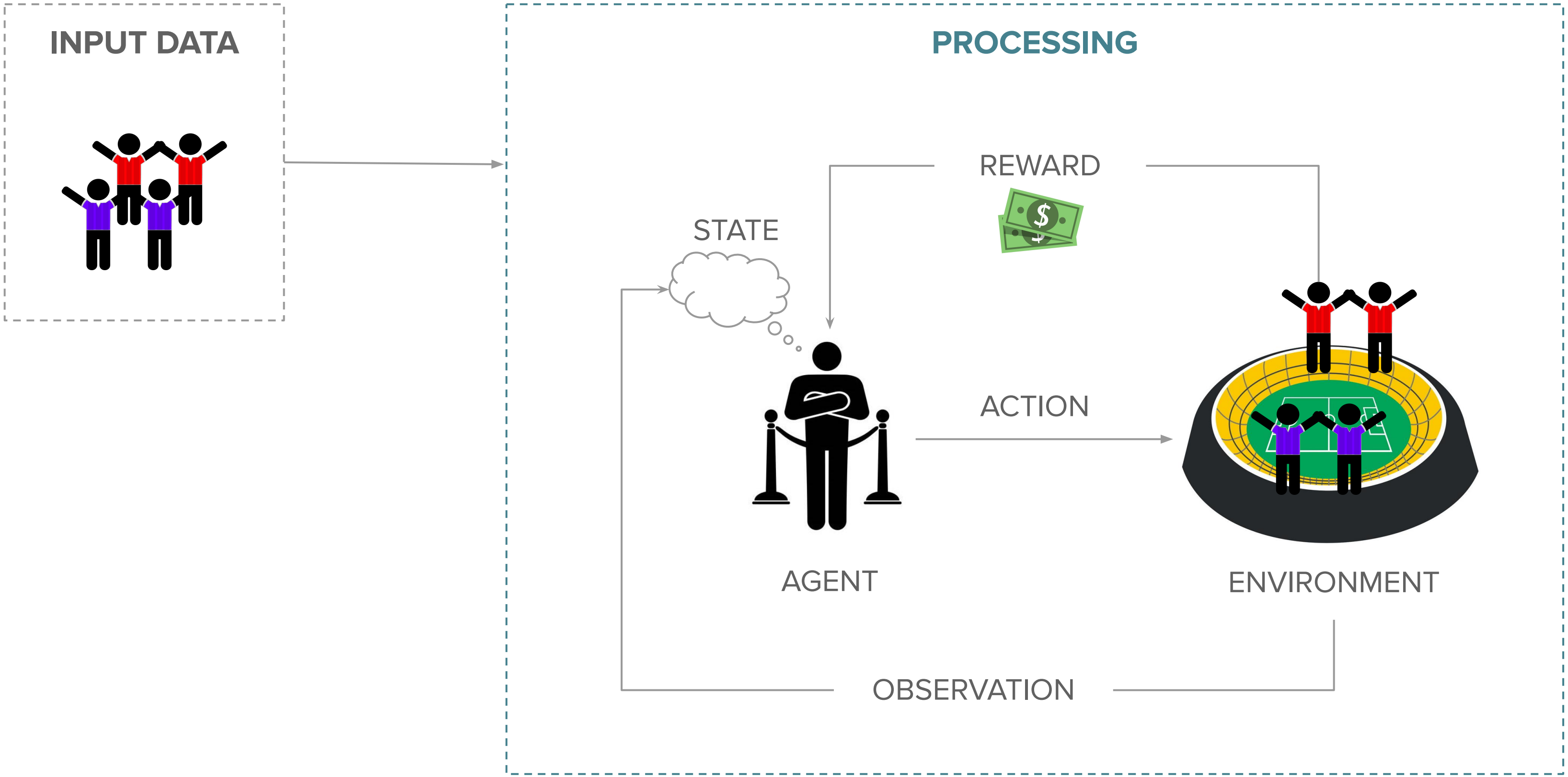
ALGORITHM

OUTPUT

SIMILAR GROUP 1

SIMILAR GROUP 2

Reinforcement Machine Learning



INPUT DATA

PROCESSING

REWARD

STATE

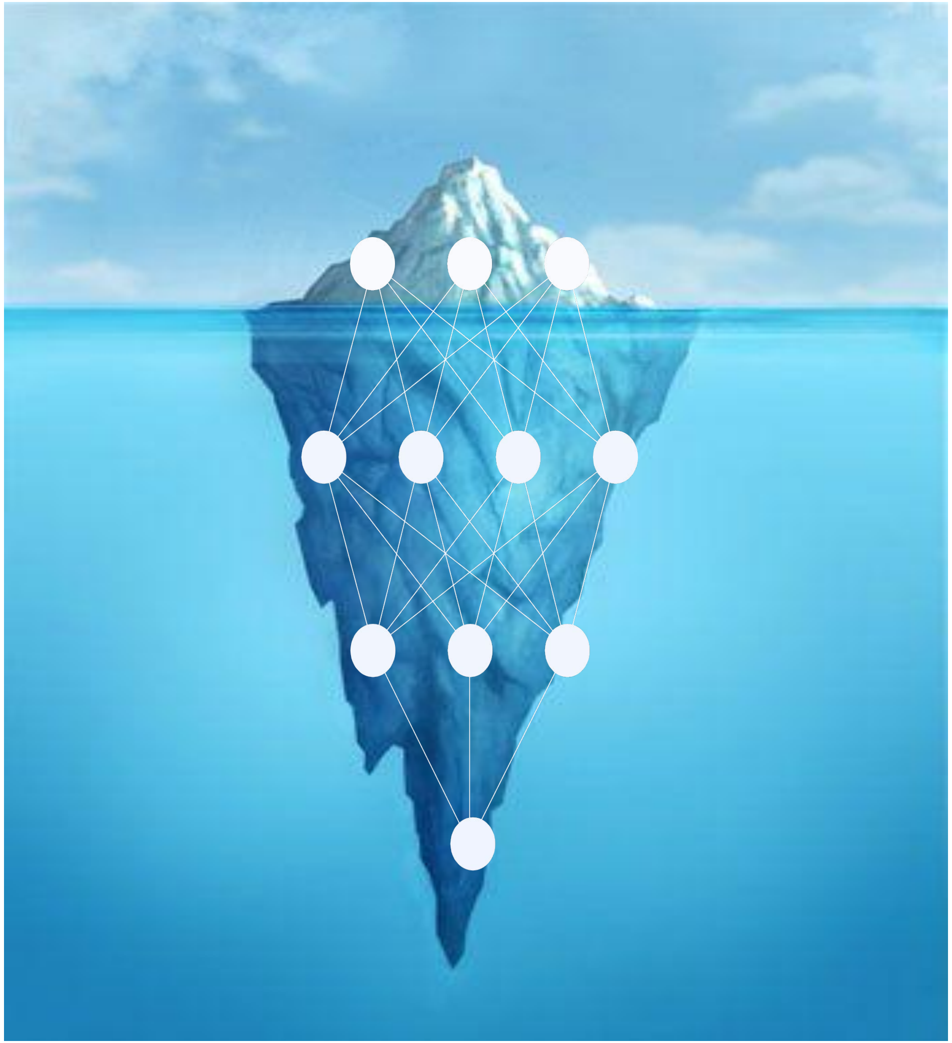
ACTION

AGENT

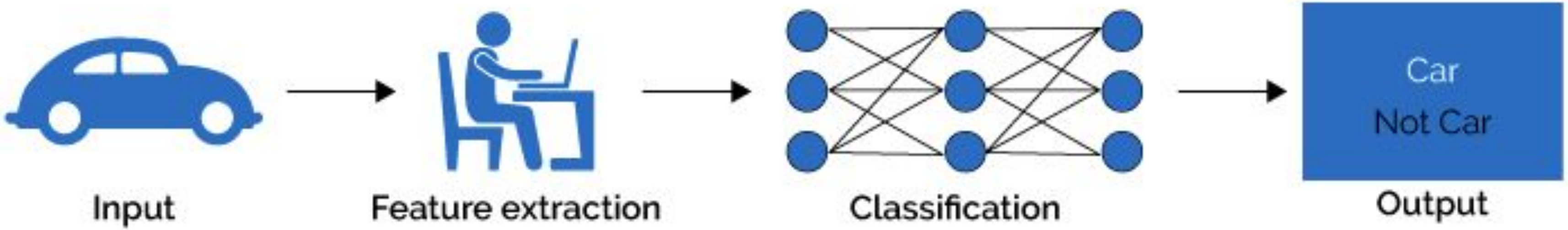
ENVIRONMENT

OBSERVATION

Deep Learning



Machine Learning



Input

Feature extraction

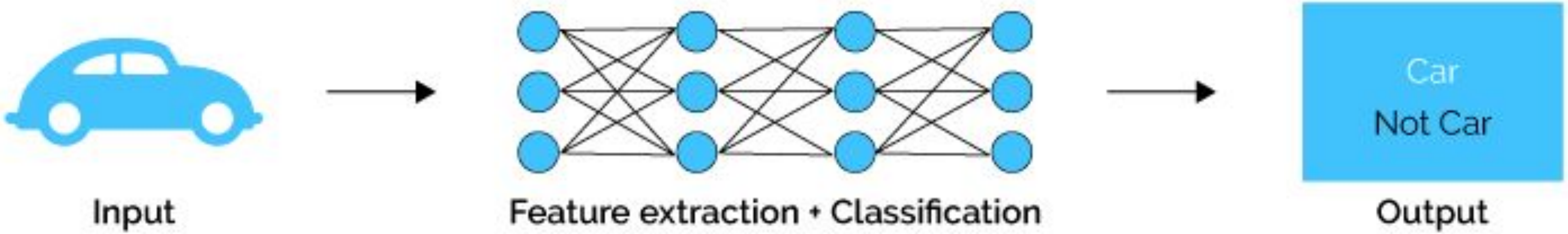
Classification

Output

Car

Not Car

Deep Learning



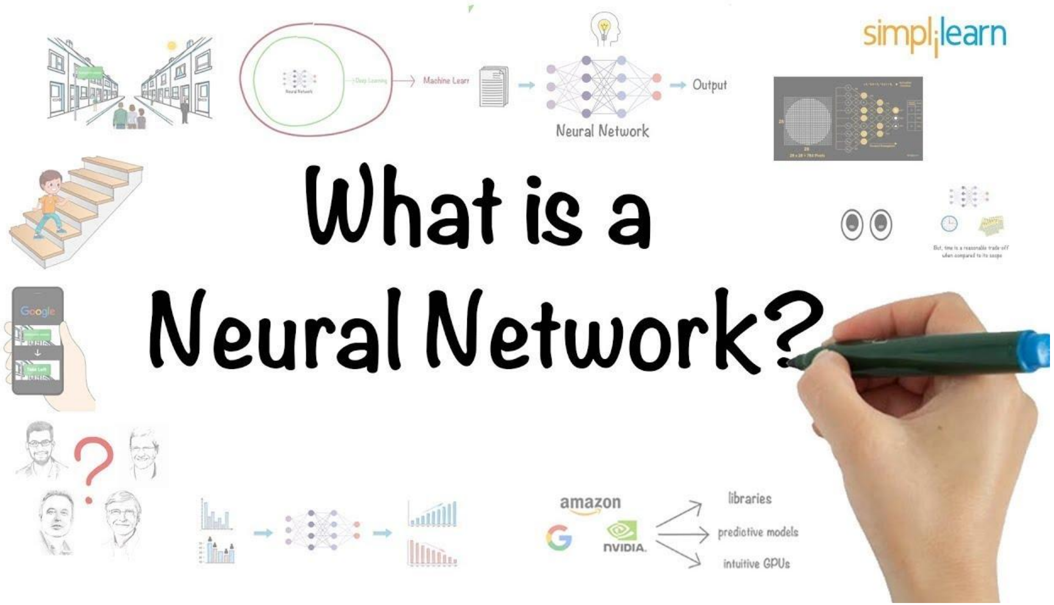
Input

Feature extraction + Classification

Output

Car

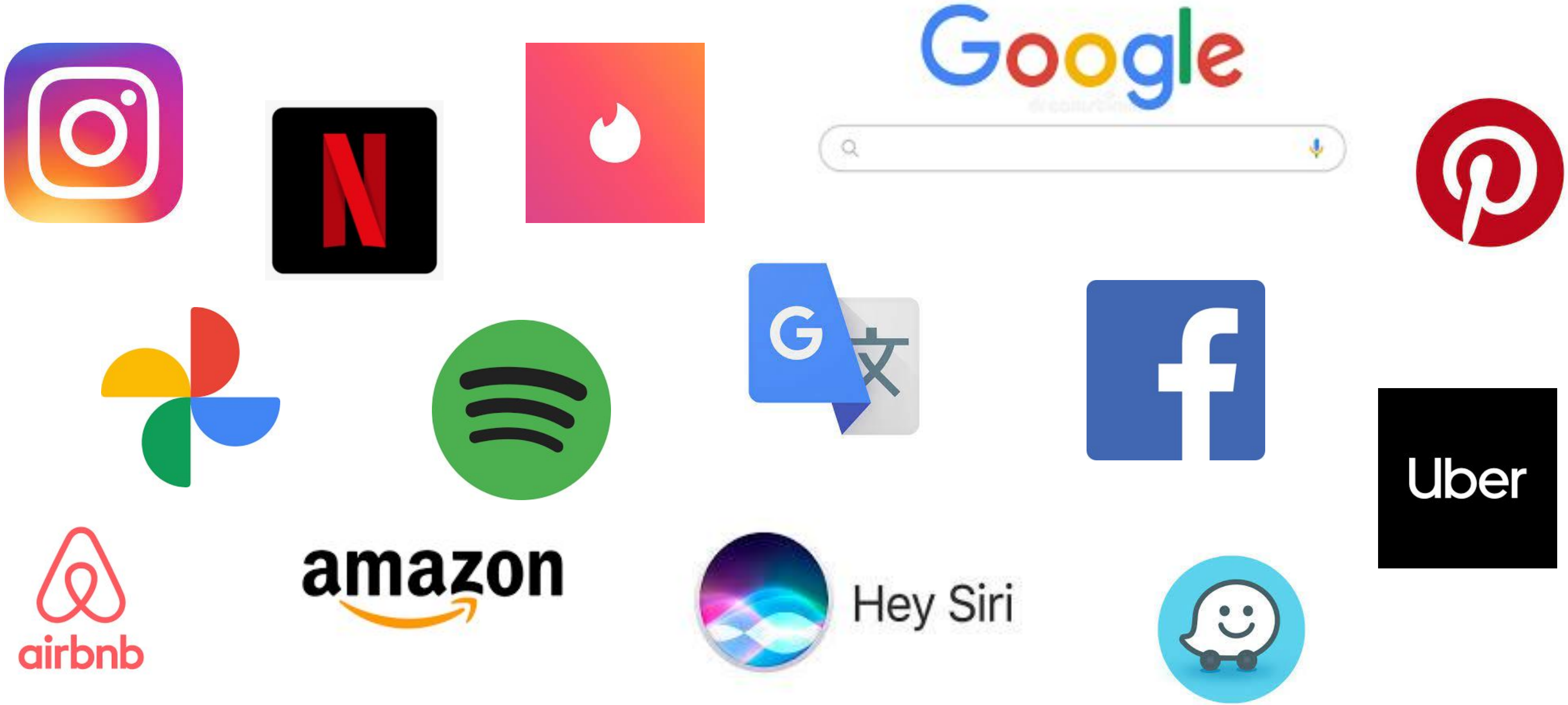
Not Car



Neural Network In 5 Minutes | What Is A Neural Network? | How Neural Networks Work | Simplilearn

https://www.youtube.com/watch?v=bfmFfD2RIcg

AI Applications

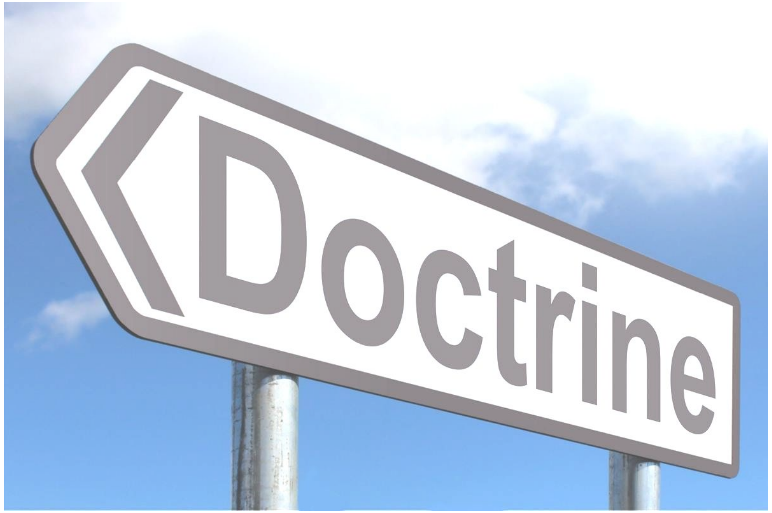

Google
N
G
文
f
Uber
airbnb
amazon
Hey Siri

Overview

- Artificial Intelligence Concepts
- AI Cyber Defense Applications
- Hands-on Exercise

Cyber Defense as “a set of **offensive**, **defensive,** and **exploratory** actions carried out in Cyberspace (...)”

- Brazilian Military Cyber Defense Doctrine, 2014

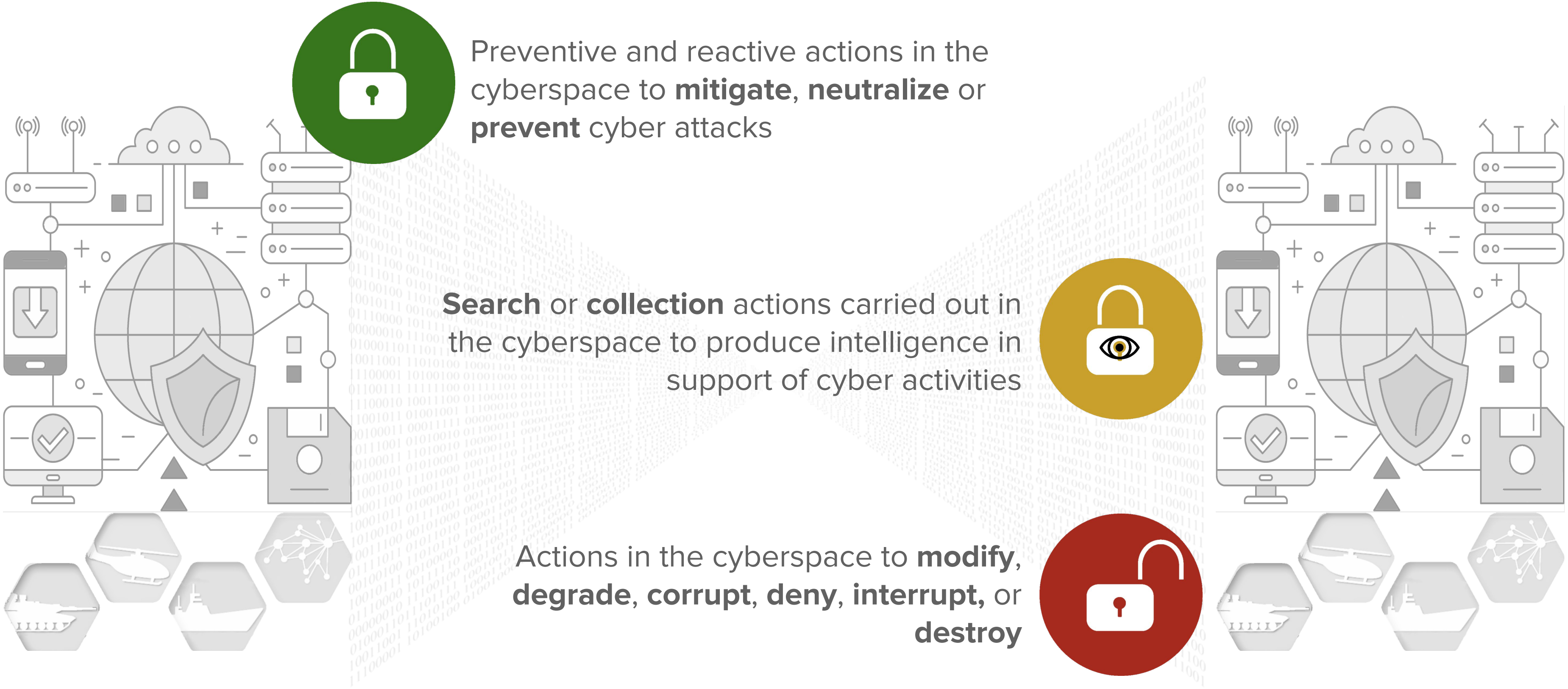


3 Types of Cyber Actions

Cyber Protection

Cyber Exploitation

Cyber Attack

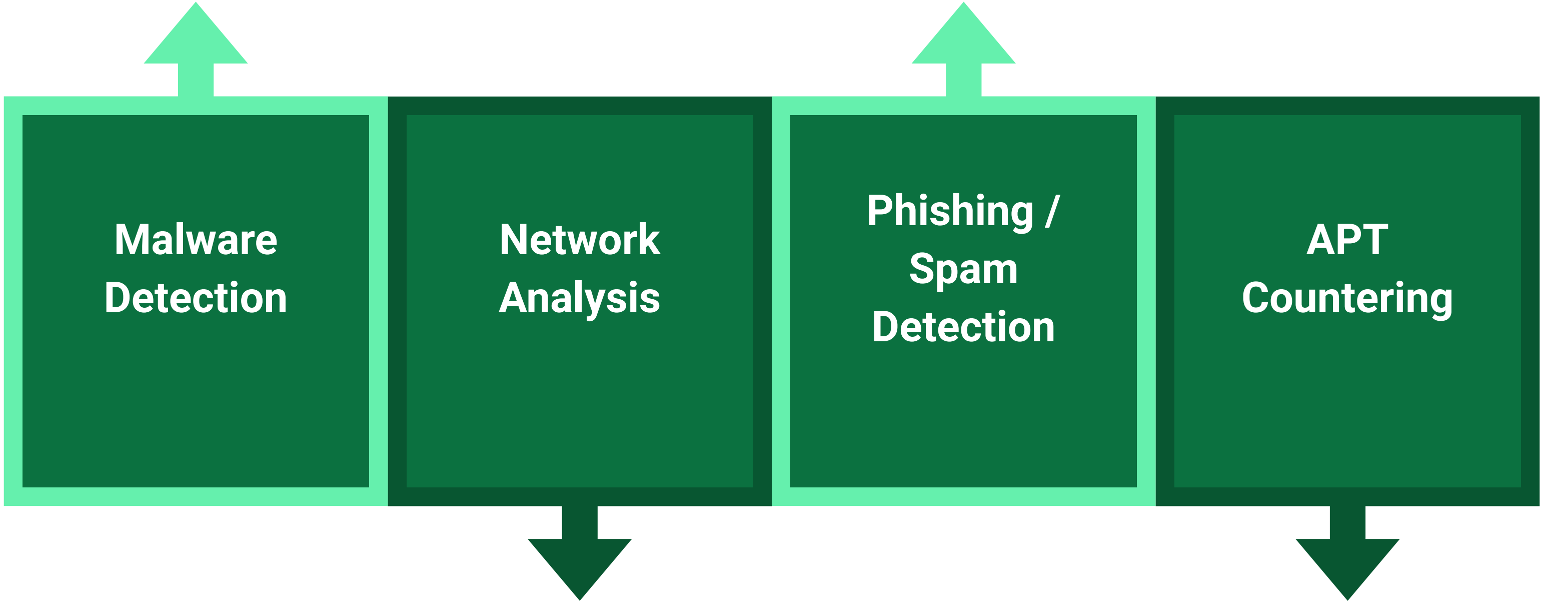


Preventive and reactive actions in the cyberspace to **mitigate**, **neutralize** or **prevent** cyber attacks

Search or **collection** actions carried out in the cyberspace to produce intelligence in support of cyber activities

Actions in the cyberspace to **modify**, **degrade**, **corrupt**, **deny**, **interrupt,** or **destroy**

AI in Cyber Protection

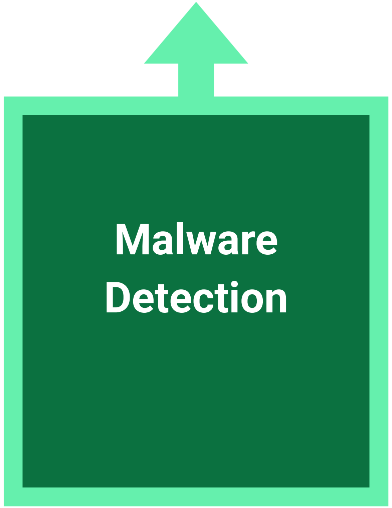


Malware Detection

Network Analysis

Phishing / Spam Detection

APT Countering



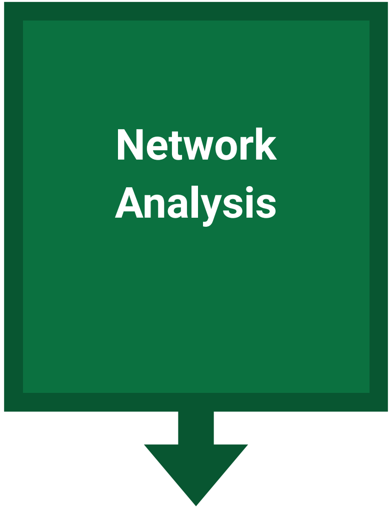
Malware Detection

Use of machine learning to classify and detect malicious software (viruses, worms, trojan horses, exploits, botnets, etc)

- Assisted detection build on base patterns
    - e.g.: Hardware utilization, virtual memory access patterns
- Performance **improvement** when compared to signature based detection
- Innovation in mobile malware classification and detection
    - e.g: ML model on mobile app permission data to distinguish between benign and malicious apps

AI to help detect threats based on application behavior and a whole network’s activity

- **Intrusion Detection**: AI-based techniques for developing and enhancing Intrusion Detection Systems (IDS), being able to outperform other techniques (higher flexibility and adaptability)
- **Anomaly Detection**: applied AI techniques to identify anomalies based on input data and previously computed network metrics



Network Analysis

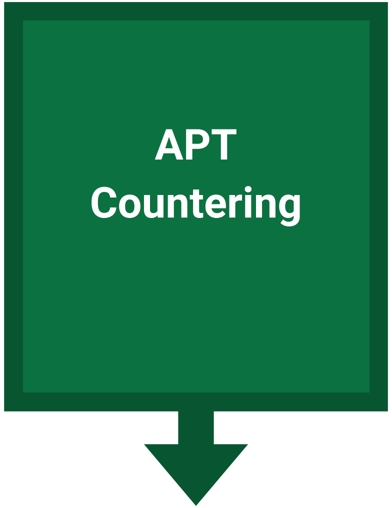
Phishing / Spam Detection

AI approaches to cope with e-mail based cyber-attacks (phishing and spam)

- Anti-phishing methods, using several different ML algorithms to distinguish **phishing websites** from legitimate ones
- Real-time anti-phishing systems based on classification algorithms and natural language processing (NLP)
- Spam ML classification models with high accuracy and efficiency

AI solutions to deal with Advanced Persistent Threats (APT)

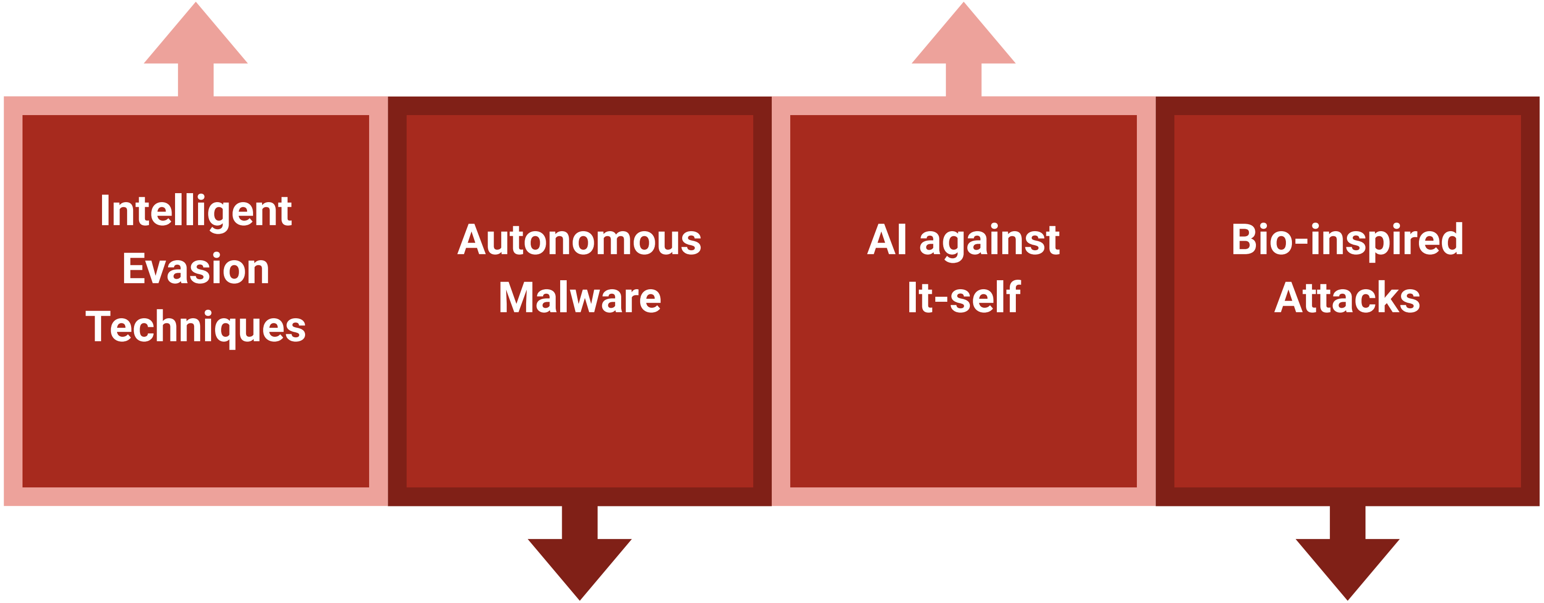
- AI enhanced IDS which can detect intrusion from the beginning of an APT to quickly react and minimize damage
- Detection of APT using machine-learning correlation analysis
    - Correlation between events from a threat detection component used as input to an attack prediction model



4 Questions to Ask a Cybersecurity AI Vendor

1. How specifically do you use AI/ML/DL in your product?
2. How was the model trained?
    a. If data set is only internal, how long is it’s minimum learning period?
3. How is the model updated?
    a. Outside data? Inside Data?
4. Is there a public case study of this feature?

AI in Cyber Exploitation and Attack



Intelligent Evasion Techniques

Autonomous Malware

AI against It-self

Bio-inspired Attacks

A.I. Is Making it Easier to Kill (You). Here’s How. | NYT

https://www.youtube.com/watch?v=GFD_Cgr2zho

Intelligent Evasion Techniques

AI as a support of one of the ultimate goals of malware: to avoid being detected by anti-malware solutions

- **Use of previous data to evade detection**: development and implementation of advanced obfuscation techniques to evade detection using data from preceding campaigns
- **Environment adjustment**: AI-powered malware that can adapt to its execution surroundings
    - Disguise as a trusted element, dodge sandbox detection
- **Use of ML for data collection**: Attackers can implement data labeling and ML to classify and capture valuable data and reduce the size of data files for stealthy exfiltration

Malware able to make calculated decisions about what to do based on its objectives and defined sensors

- **Smarter malware**: Malware could propagate based on a sequence of autonomous decisions, intelligently custom-made to the parameters of the host system
- **Eliminate C2 channels**: Malware could be equipped with intelligent automation and preliminary logical process to automatically navigate a compromised network, select the desired target, and push data to the malware owner

Autonomous Malware

AI against It-self

As AI is being integrated into security solutions, an attacker might attempt to hijack it by any means

- **AI in adversarial attacks**: a rising trend of AI-based threats, where malicious actors design the inputs to make models predict erroneously
- **Poisoning data**: Poisoning of anti-malware engines input data, so that the ML output is poisoned
    - An attacker could pollute the training data from which the algorithm is learning in such a way that the system misbehaves

Algorithms inspired by nature

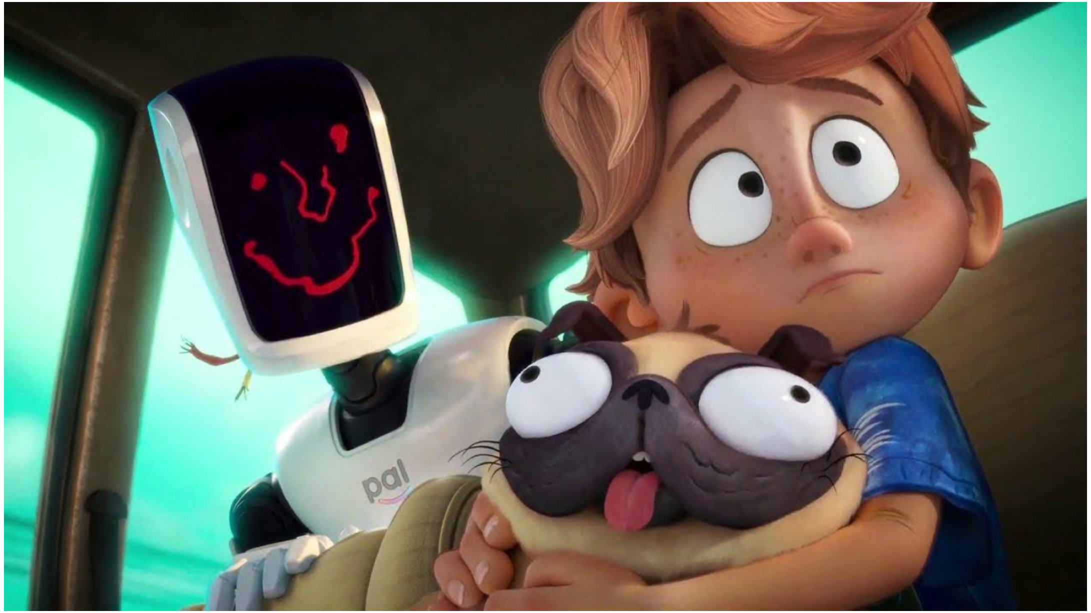
- **Mutant malware**: Creation of obfuscated malware, capable of overcoming modern detection tools
  - e.g.: A malware that rewrites its code structure whenever it is executed
- **Swarm-based intelligence malware**: An option for surpassing a C2 centralized structure
  - e.g.: A framework of swarm intelligence-based algorithms could enable decentralized communication, using the simulated behavior of biological swarm systems and creating a “collective memory”

Bio-inspired Attacks

Adversarial Machine Learning: What? So What? Now What?

https://www.youtube.com/watch?v=JsklJW01bjc

[2], [3]



Dog...? Pig...? (From "The Mitchells vs The Machines" Movie)

https://www.youtube.com/watch?v=7T06Mml2-jE

[2], [3]

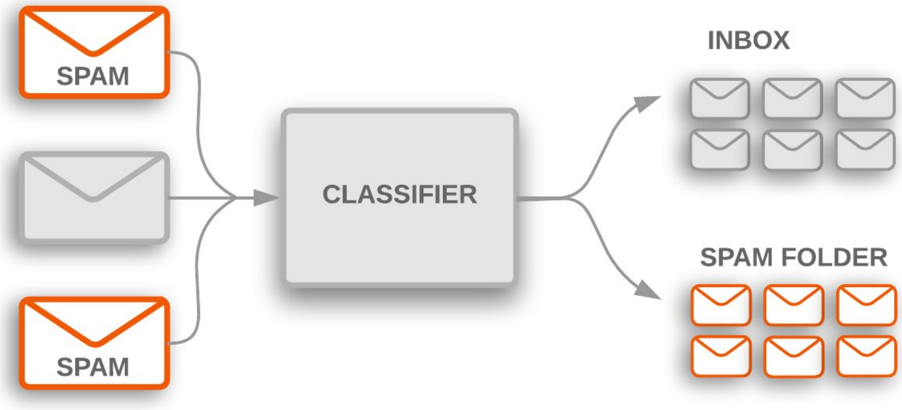
Overview

- Artificial Intelligence Concepts
- AI Cyber Defense Applications
- Hands-on Exercise

SCENARIO

- Mail admin
- Needs to automatically flag **phishing** messages as spam

Supervised Machine Learning



Two Exercises:

1. Feature based classification
2. Natural language processing

Data sets

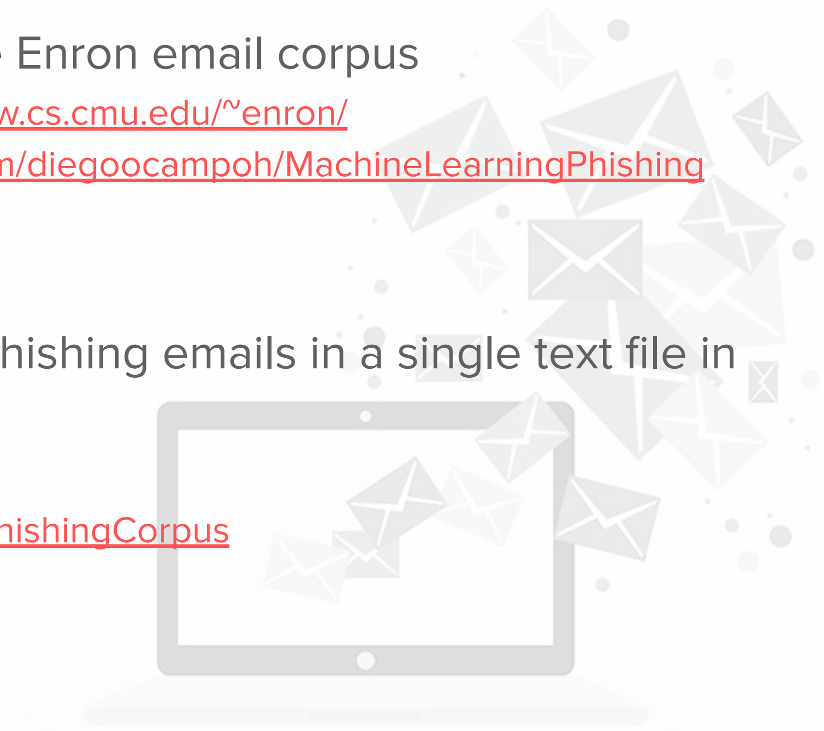
A. 2000 regular, non-phishy emails from the Enron email corpus

   a. William W. Cohen, MLD, CMU, 2015. https://www.cs.cmu.edu/~enron/

   b. Converted to mbox format by https://github.com/diegoocampoh/MachineLearningPhishing

B. Phishing email corpus containing 2000 phishing emails in a single text file in the mbox format

   a. Nazario. phishingcorpus homepage, 2006. http://monkey.org/%7Ejose/wiki/doku.php?id=PhishingCorpus

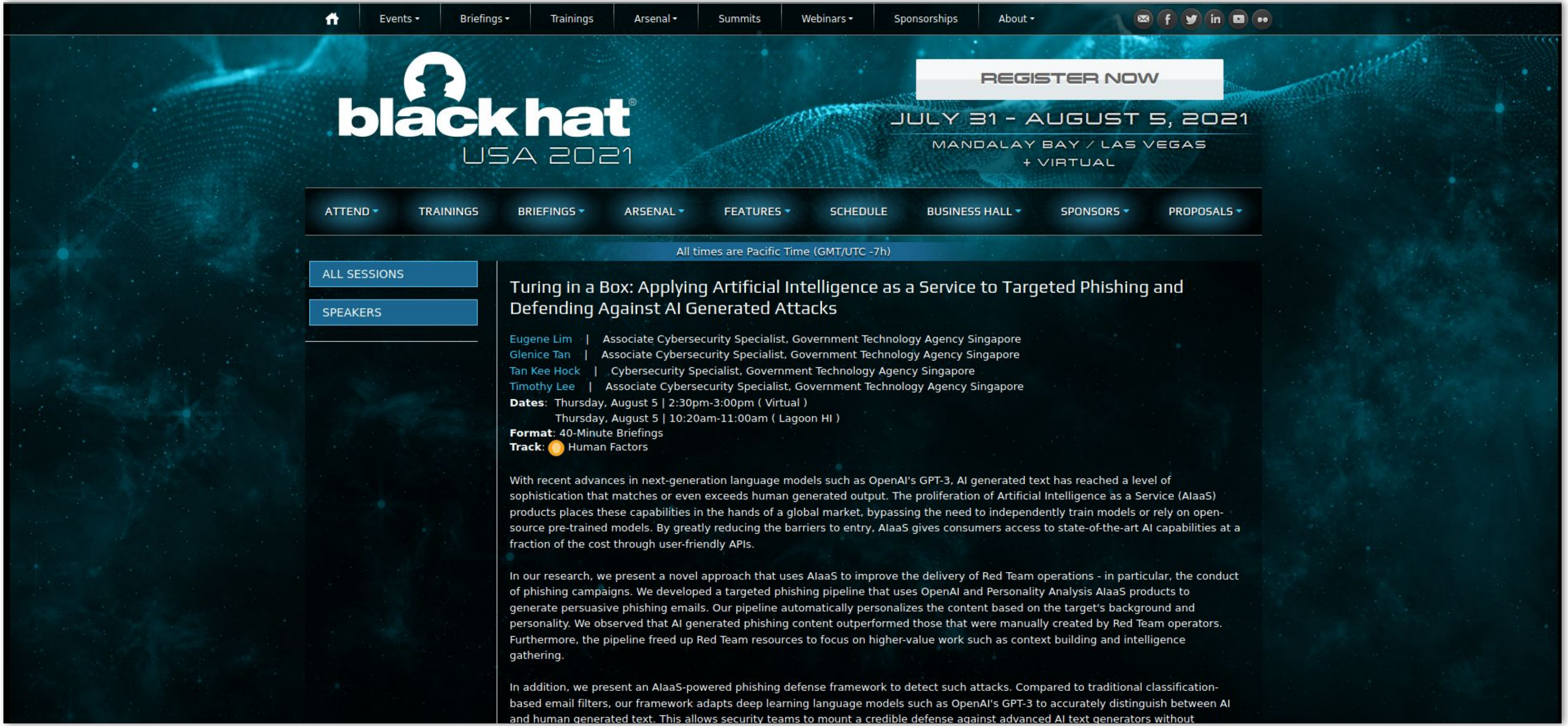


Tools

- Jupyter notebook
    - Free, open-source, interactive web web interface to **Python**
    - Combines software code, computational output, explanatory text and multimedia resources in a single document
- Scikit-learn
    - Free software machine learning library for **Python**
- Docker
    - Open source containerization platform
    - Package applications into containers
        - jupyter/scipy-notebook

Docker

```
 #    docker images

REPOSITORY              TAG       IMAGE ID        CREATED        SIZE
jupyter/scipy-notebook   latest    9e1ff2e82f6a   X days ago   2.58GB



 #    apt update && apt install -y docker.io
 #    systemctl enable docker --now
 #    docker pull jupyter/scipy-notebook



 #    git clone https://github.com/isabellecda/supervised-ml-training.git
 #    cd supervised-ml-training
 #    docker run -p 8888:8888 -v $(pwd):/home/jovyan/work jupyter/scipy-notebook
```



Turing in a Box: Applying Artificial Intelligence as a Service to Targeted Phishing and Defending Against AI Generated Attacks

Eugene Lim | Associate Cybersecurity Specialist, Government Technology Agency Singapore
Glenice Tan | Associate Cybersecurity Specialist, Government Technology Agency Singapore
Tan Kee Hock | Cybersecurity Specialist, Government Technology Agency Singapore
Timothy Lee | Associate Cybersecurity Specialist, Government Technology Agency Singapore

Dates: Thursday, August 5 | 2:30pm-3:00pm (Virtual)
Thursday, August 5 | 10:20am-11:00am (Lagoon HI)
Format: 40-Minute Briefings
Track: ● Human Factors

With recent advances in next-generation language models such as OpenAI's GPT-3, AI generated text has reached a level of sophistication that matches or even exceeds human generated output. The proliferation of Artificial Intelligence as a Service (AIaaS) products places these capabilities in the hands of a global market, bypassing the need to independently train models or rely on open-source pre-trained models. By greatly reducing the barriers to entry, AIaaS gives consumers access to state-of-the-art AI capabilities at a fraction of the cost through user-friendly APIs.

In our research, we present a novel approach that uses AIaaS to improve the delivery of Red Team operations - in particular, the conduct of phishing campaigns. We developed a targeted phishing pipeline that uses OpenAI and Personality Analysis AIaaS products to generate persuasive phishing emails. Our pipeline automatically personalizes the content based on the target's background and personality. We observed that AI generated phishing content outperformed those that were manually created by Red Team operators. Furthermore, the pipeline freed up Red Team resources to focus on higher-value work such as context building and intelligence gathering.

In addition, we present an AIaaS-powered phishing defense framework to detect such attacks. Compared to traditional classification-based email filters, our framework adapts deep learning language models such as OpenAI's GPT-3 to accurately distinguish between AI and human generated text. This allows security teams to mount a credible defense against advanced AI text generators without

https://www.blackhat.com/us-21/briefings/schedule/#turing-in-a-box-applying-artificial-intelligence-as-a-service-to-targeted-phishing-and-defending-against-ai-generated-attacks-22925

References

[1] When an AI Goes Full Jack Kerouac. Brian Merchant. The Atlantic, 2018.
https://www.theatlantic.com/technology/archive/2018/10/automated-on-the-road/571345/

[2] Turing Test Illustration. TECHTARGET, 2017.
https://searchenterpriseai.techtarget.com/definition/Turing-test

[3] Computing Machinery and Intelligence. Alan Turing. Mind, 1950. https://doi.org/10.1093/mind/LIX.236.433

[4] Artificial Intelligence: What is Artificial Intelligence? How Does AI Work? Tech Hiring Company. 2021.
https://builtin.com/artificial-intelligence

[5] Doutrina Militar de Defesa Cibernética. Ministério da Defesa, 2014.
https://bdex.eb.mil.br/jspui/handle/123456789/136

References (2)

[6] Artificial Intelligence in the Cyber Domain: Offense and Defense. Truong, Diep and Zelinka. Symmetry, 2020. https://doi.org/10.3390/sym12030410

[7] Machine Learning Fundamentals for Cybersecurity Professionals. Vectra, 2020. https://www.vectra.ai/resources/vid-machine-learning-fundamentals-for-cybersecurity-professionals

[8] Vendor Data Science Buzzwords Hacked. Thordis Thorsteins. 2019. https://www.youtube.com/watch?v=IXm5gp0ZGWM

[9] A Survey on Artificial Intelligence in Malware as Next-Generation Threats. Truong and Zelinka. Mendel, 2019. https://doi.org/10.13164/mendel.2019.2.027

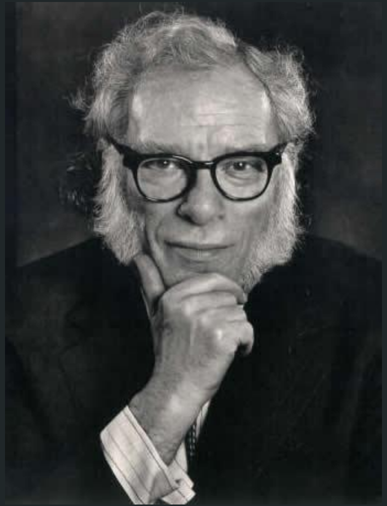
Overview

- Artificial Intelligence Concepts
- AI Cyber Defense Applications
- Hands-on Exercise

Objectives

- To get to know the main concepts of Artificial Intelligence and its applications to the cyber domain
- To be able to apply these concepts in a simulated scenario

*“In a properly automated and educated world, then, machines may prove to be **the true humanizing influence**. It may be that machines will do the work that makes life possible and that human beings will do all the other things that make life **pleasant and worthwhile**”*



– Isaac Asimov, Writer and Professor, on ‘Robot Visions’