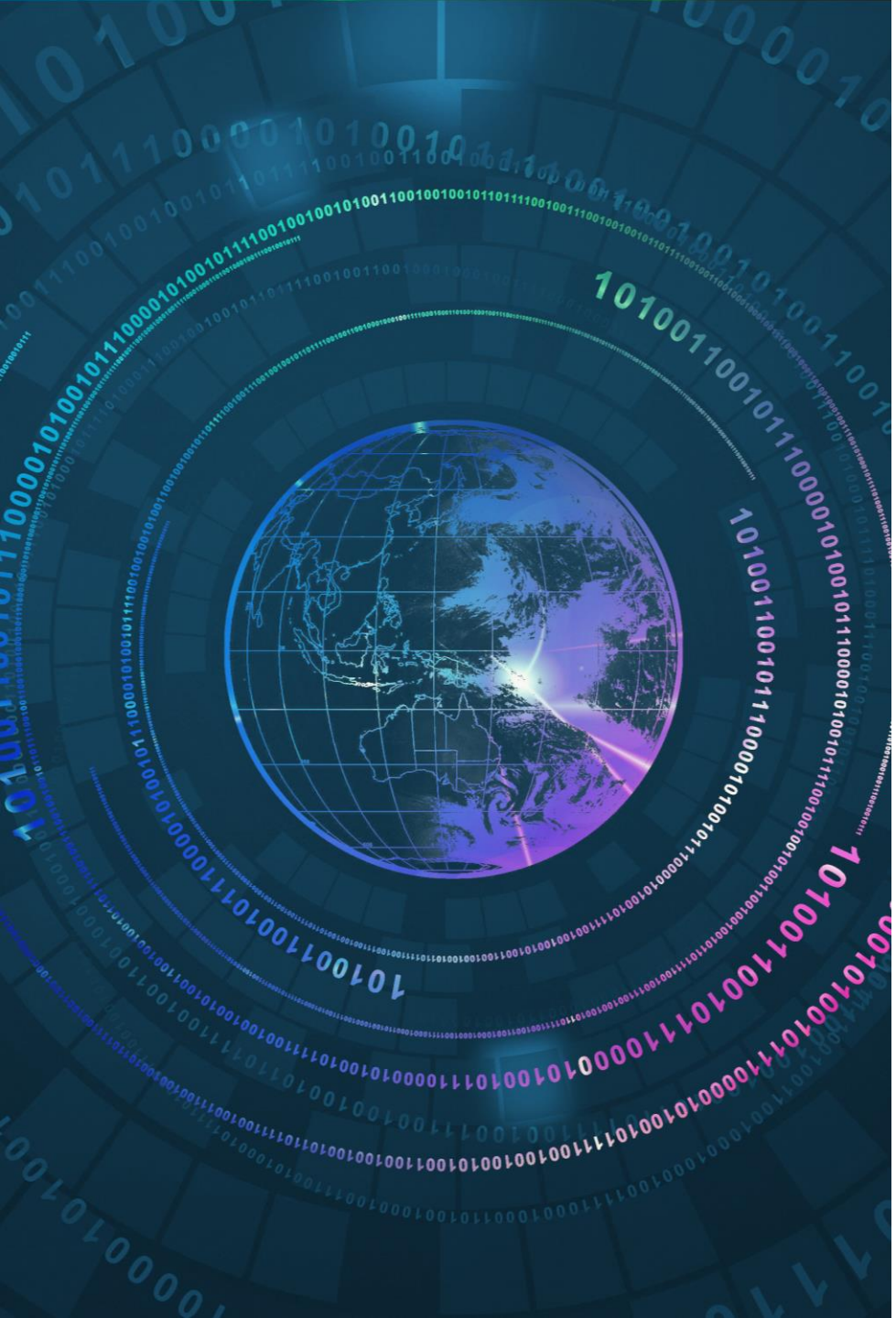


Roles of Artificial Intelligence and Deep Learning in Detecting Data Exfiltration

PARALLAX CYBER LLC

My Background

Chief Technologist, Cyber Operations
L3Harris Technologies, Space & Airborne Systems

President: Parallax Cyber LLC

Education:
B.S. in Computer Science, Angelo State University, TX
Software Professional Development Program,
Air Force Institute of Technology, OH
M.S. in Computer Information Systems, Regis University, CO
D.Sc. in Cybersecurity, Capitol Technology U., MD

Member: ISSA, IEEE Computer Society, INCOSE, ISC2, SANS

Certifications:
CISSP (ISC2)
Certified Ethical Hacker (EC-Council)
GIAC Reverse Engineering Malware (SANS)
Certified Systems Engineering Professional (INCOSE)

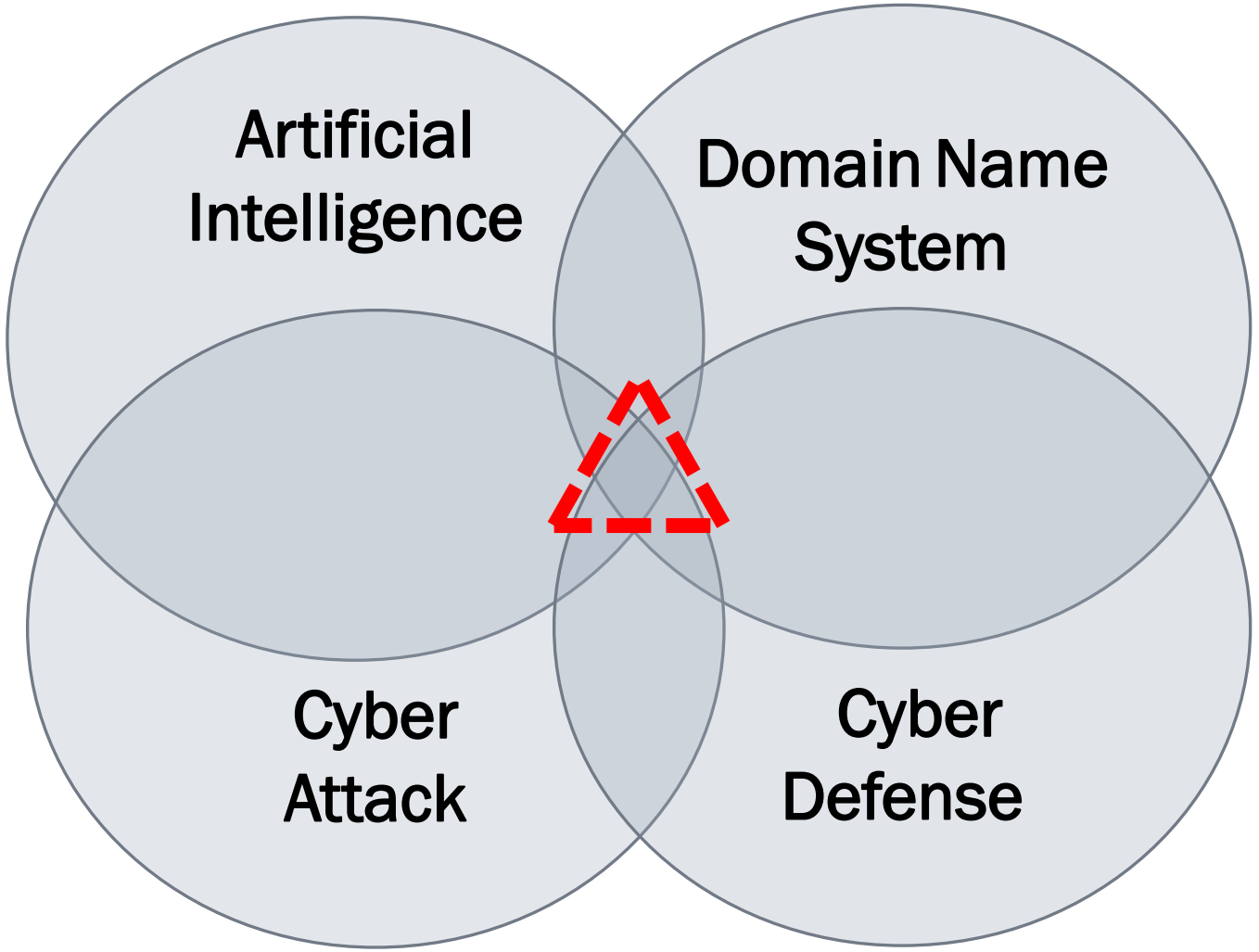
Contact: tomas.pena@[parallax-cyber.com]

Opinions expressed are my own and do not represent the opinions of L3Harris Technologies.

Sources – All are “Open Source”

See REFERENCES section

Multi-Discipline Approach



Artificial Intelligence

Able to recognize patterns in data and calculate proximity to pattern

Originated in 1956, popularity ebbed and flowed due to overpromising

Advanced in modern technologies have spurred growth:

- Algorithm & Tool Development (Tensorflow, Pytorch, Anaconda Distro, etc.)
- Processing power (Graphical Processor Units – GPUs) & Distributed Processing
- Data science (big data)

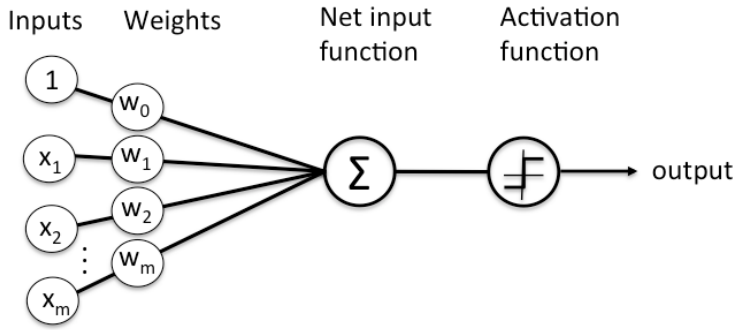
Major categories recognized

- General AI – Able to discover data, Self-Aware
- Narrow AI – Machine Learning (ML)

Machine Learning

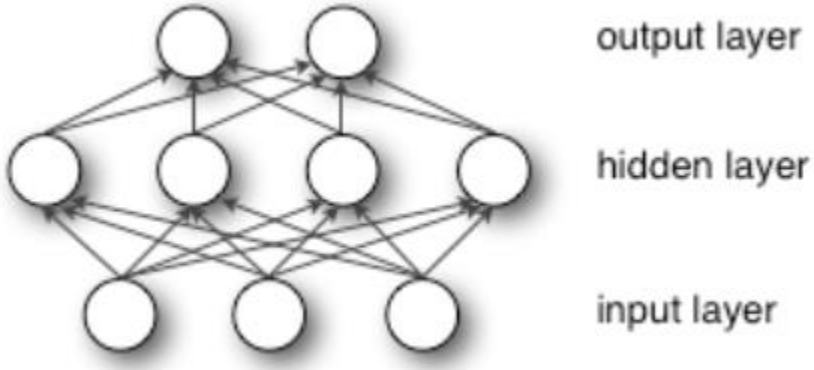
Shallow Learning Algorithms

- Simple gradient-based algorithms
- Suitable for low dimension data (few features to consider)



Deep Learning Algorithms

- Complex gradients
- High-dimensionality (facial recognition, image recognition, etc.)
- Computation intensive
- Complex neural networks &



Images: https://wiki.pathmind.com/neural-network

Machine Learning

Why Deep Learning?

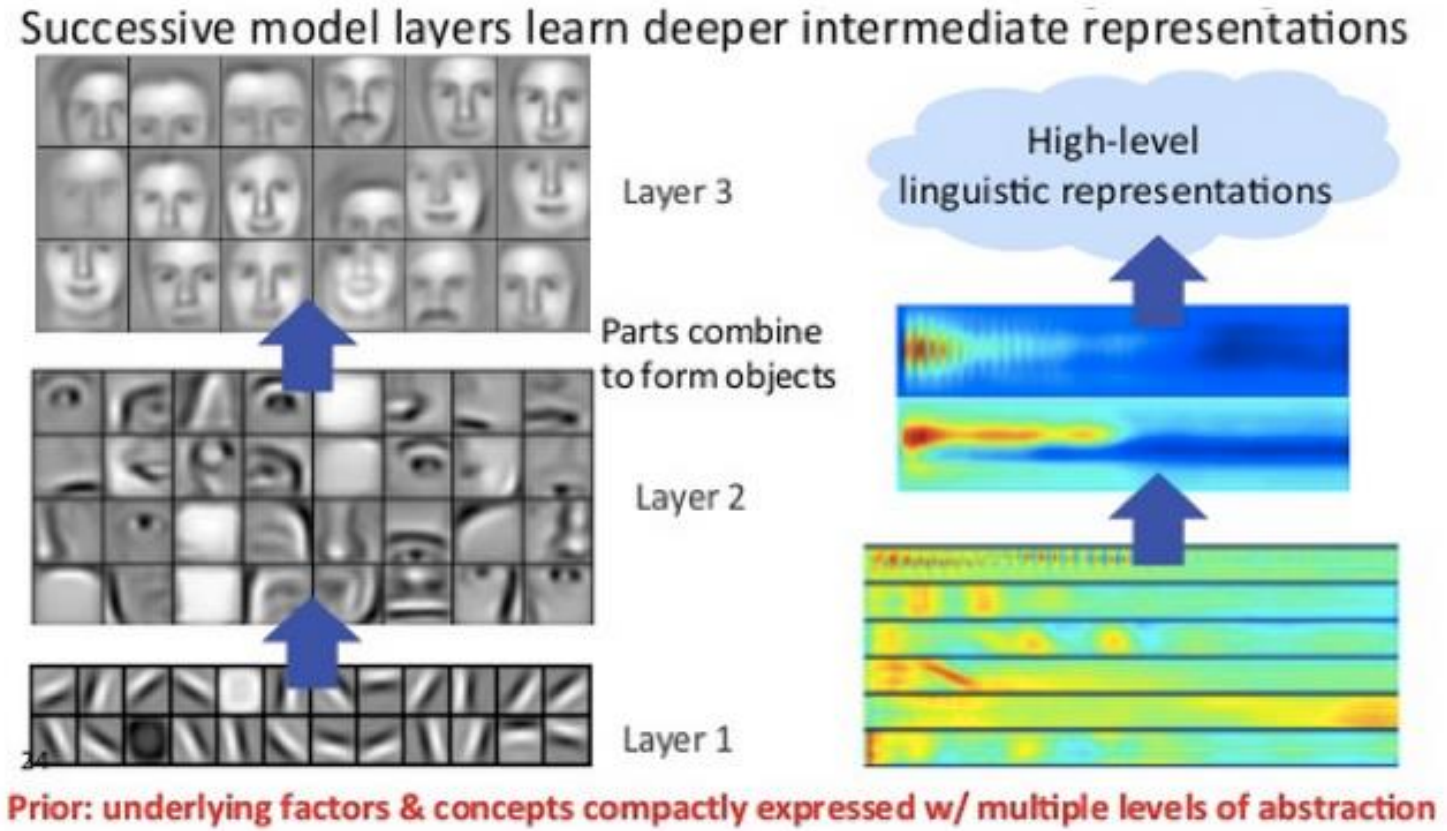


Image: https://wiki.pathmind.com/neural-network

DNS

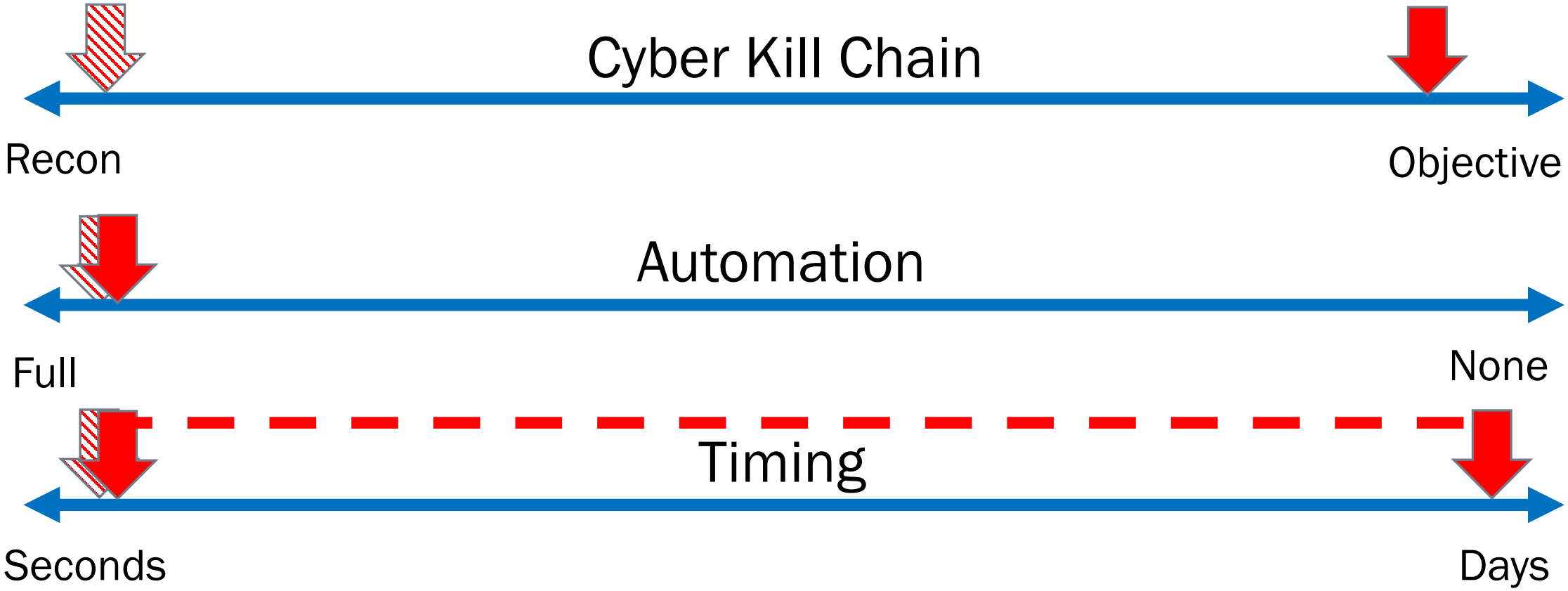
Essential Internet service designed to locate resources since 1983.

Standard configuration and well-known protocols in place (UDP, Port 53).

Highly vulnerable system subject to multiple forms of exploitation

- Direct users to malicious websites to download malware (Domain hijacking)
- Man-in-th-Middle attacks using MX record and custom PKI certificate
- DNS over HTTPS to encrypt malicious traffic
- Data exfiltration over open port

Detection & Alerting



>DNSExfiltrator

Open Source Program (Server/Client)

Encodes data into fragments of Base-32/-64; transmits in UDP packets

Leverages client network DNS configuration

- Default configuration to reach out for domain (ala Domain Generation Algorithm)
- DNS formatted-traffic to designated IP address using DNS UDP format and ports

Multiple options for obfuscating exfiltration

- Variable packet size
- Variable packet timing
- Encryption

Example of data exfiltration:

`148.mnROV8y9hzMJn1K6JQ3Qi67B4uosWjBmAUo.CapTechUtest.org`

https://github.com/Arno0x/DNSExfiltrator

Experiment

Generated 300+ files of different sizes

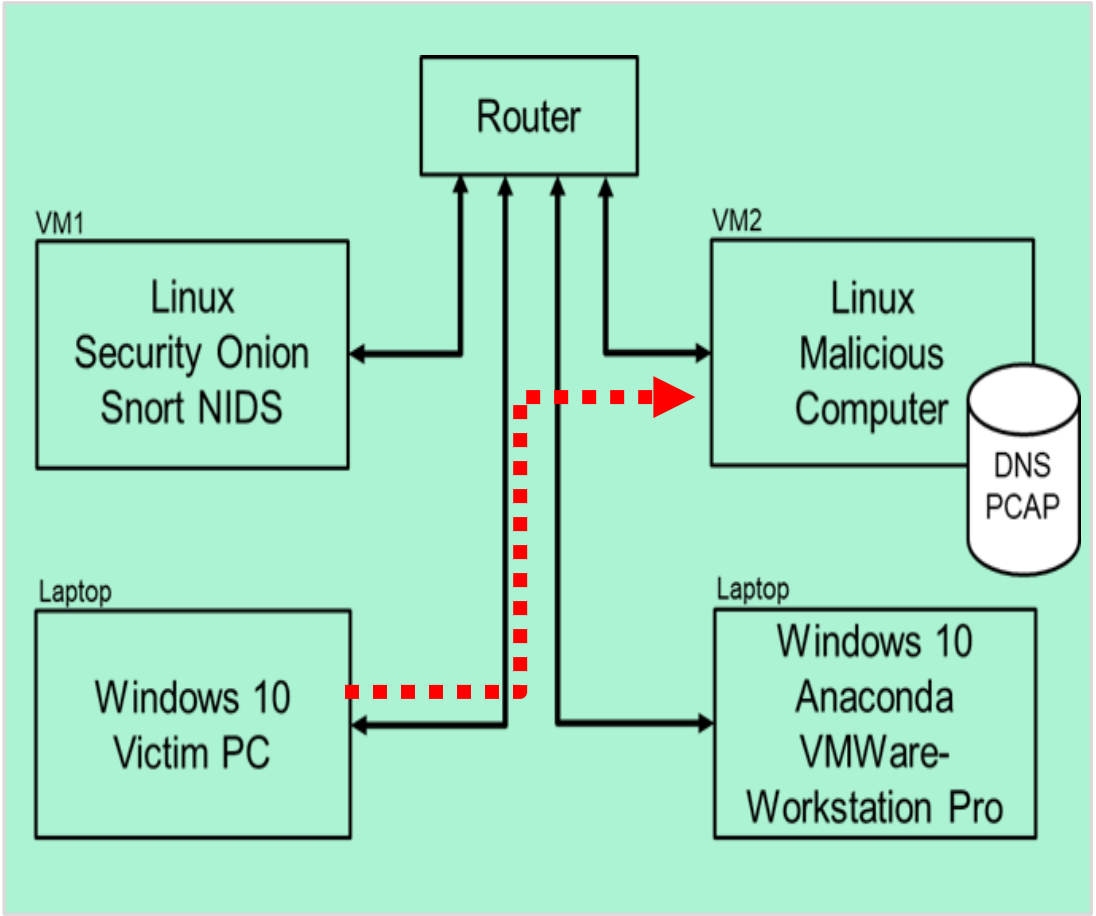
Examined DNSExfiltrator parameter effects

- Packet timing, encryption
- Packet size (MXSIZE)

Captured network traffic (PCAP)

Built DL model (named Fedona)

- Convolutional Neural Network (1D)
- Cisco Umbrella DNS data (Known good)
- DNXExfiltration traffic capture (Known bad)



Local Lab Configuration

ML Model Architecture

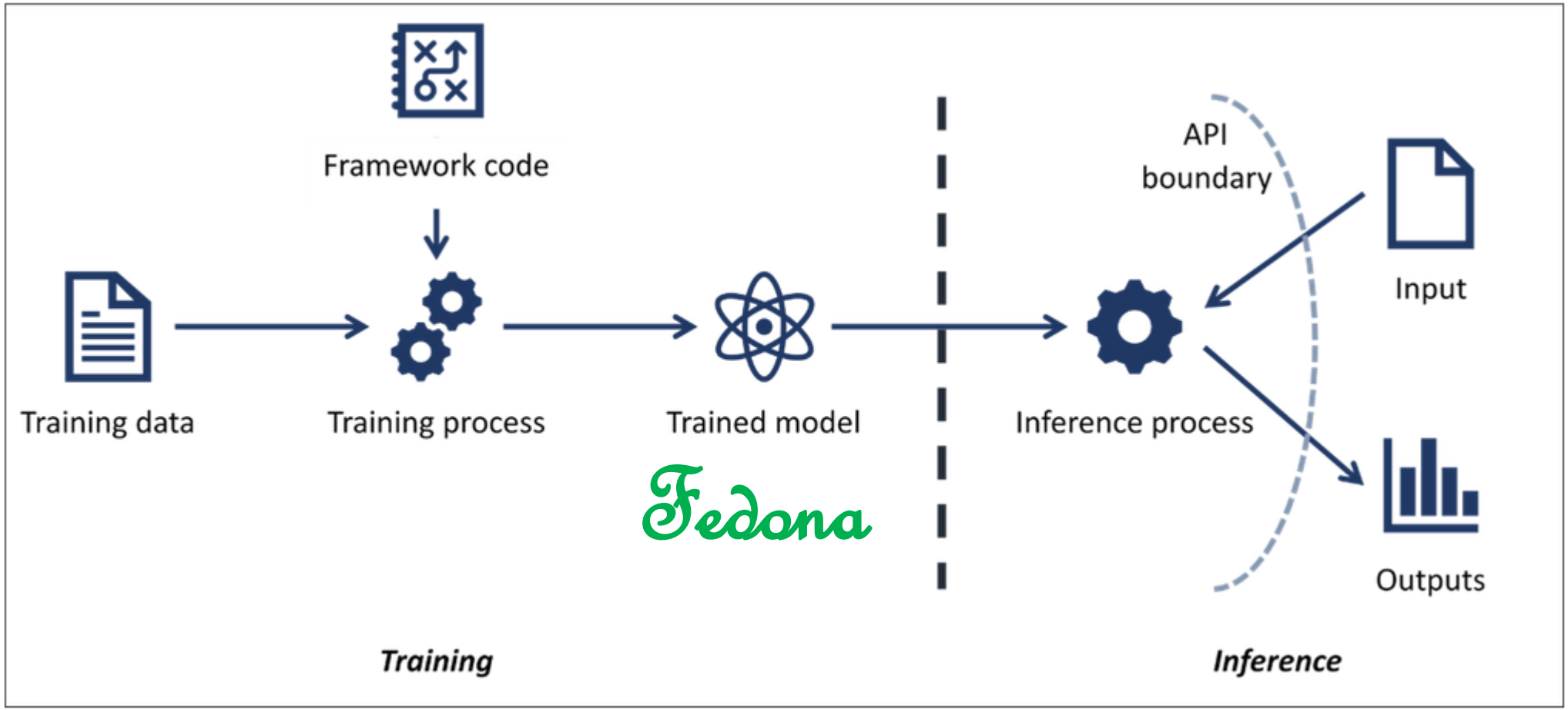
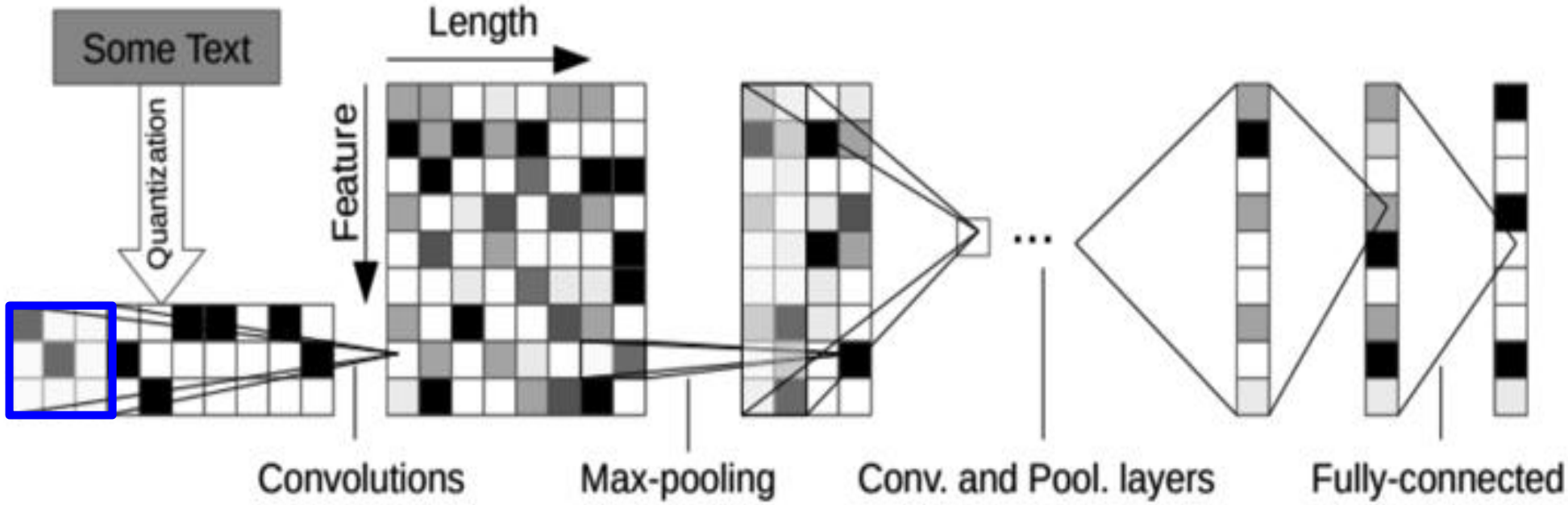


Image: https://github.com/mitre/advmlthreatmatrix/blob/master/images/AdvML101.PNG

Convolutional Neural Network (CNN) for Natural Language Processing (NLP)

Base64 encoding characters = [A..Za..z0..9+/=]

148.mnROV8y9hzMJn1K6JQ3Qi67B4uosWjBmAUo.CapTechUtest.org



Results

Zero False Positives

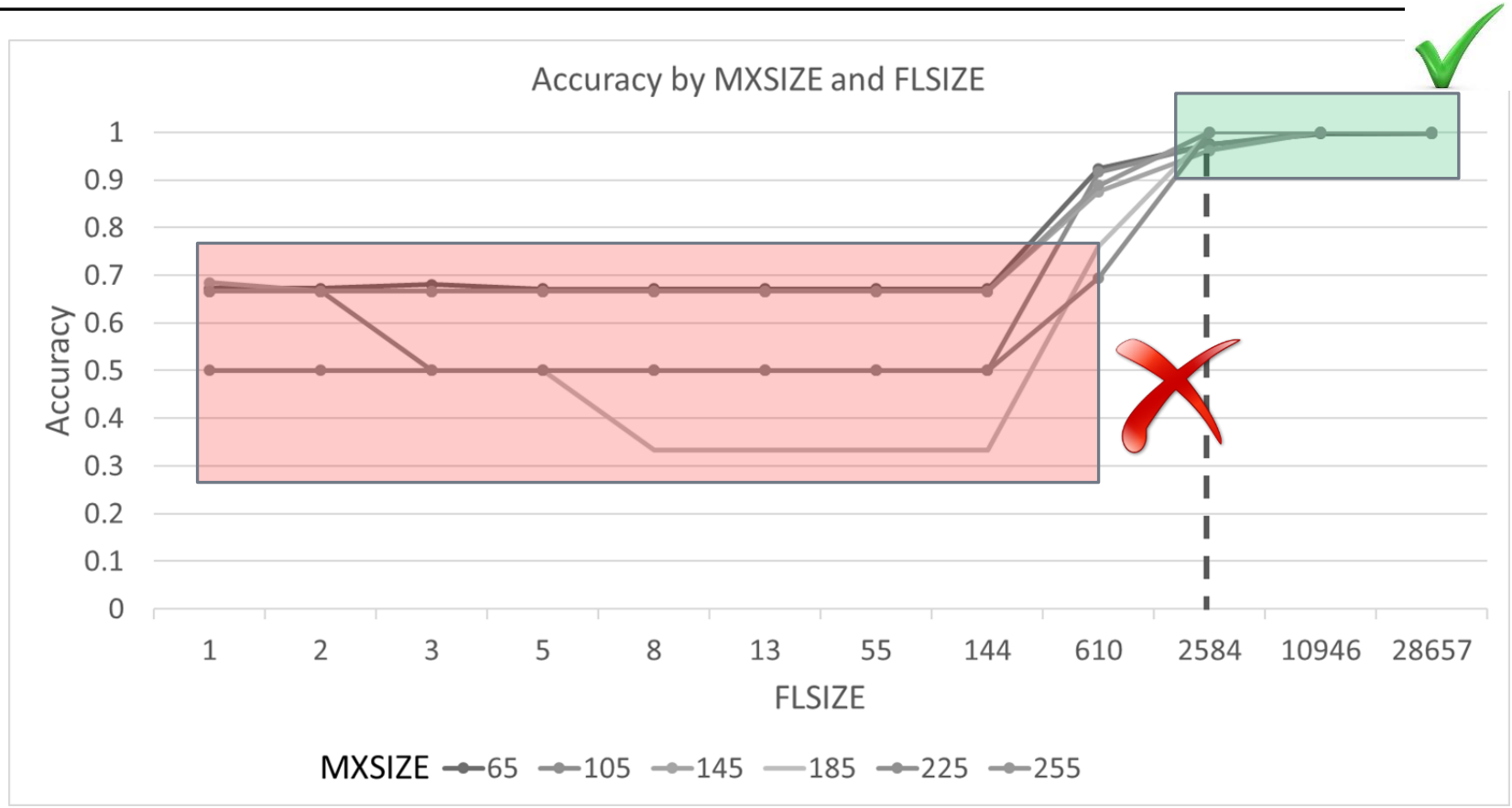
- Possibly “over fitted”

Highly accurate

- Default DNSExfiltrator parameters

Poor accuracy

- Smaller FLSIZE
- Smaller MXSIZE



AI Applied to Cybersecurity

Defensive

- Detection
- Response

Offensive

- Target identification
- Defense detection & defeat
- Attack ML systems

According to a Gartner report. 30% of cyberattacks by 2022 will involve data poisoning, model theft or adversarial examples.

https://www.gartner.com/en/documents/3939991

Adversarial ML Threat Matrix – Introduced October 2020
https://github.com/mitre/advmlthreatmatrix

ML Model Architecture

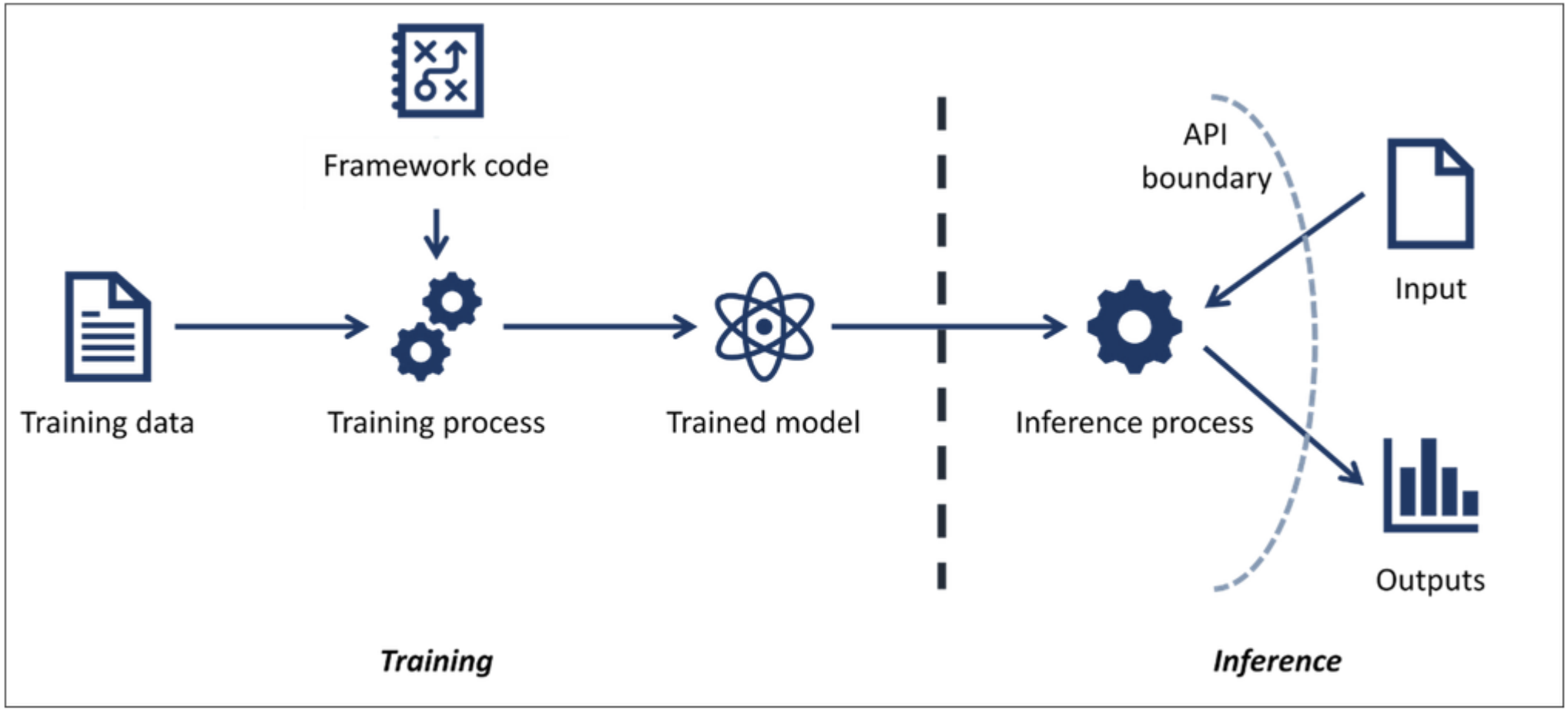


Image: https://github.com/mitre/advmlthreatmatrix/blob/master/images/AdvML101.PNG

Adversarial Machine Learning (ML) Threat Matrix

ML Attacks Types

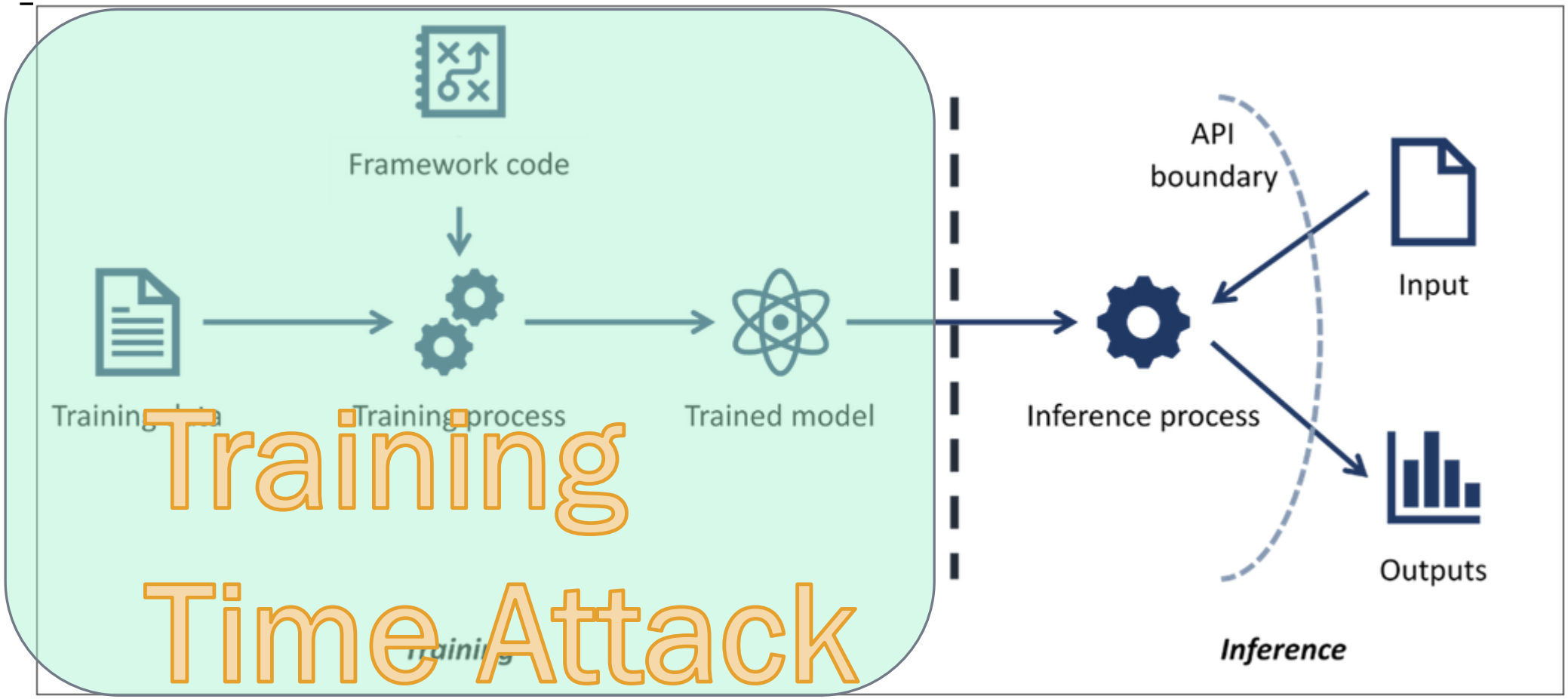


Image: https://github.com/mitre/advmlthreatmatrix/blob/master/images/AdvML101.PNG

ML Attacks Types

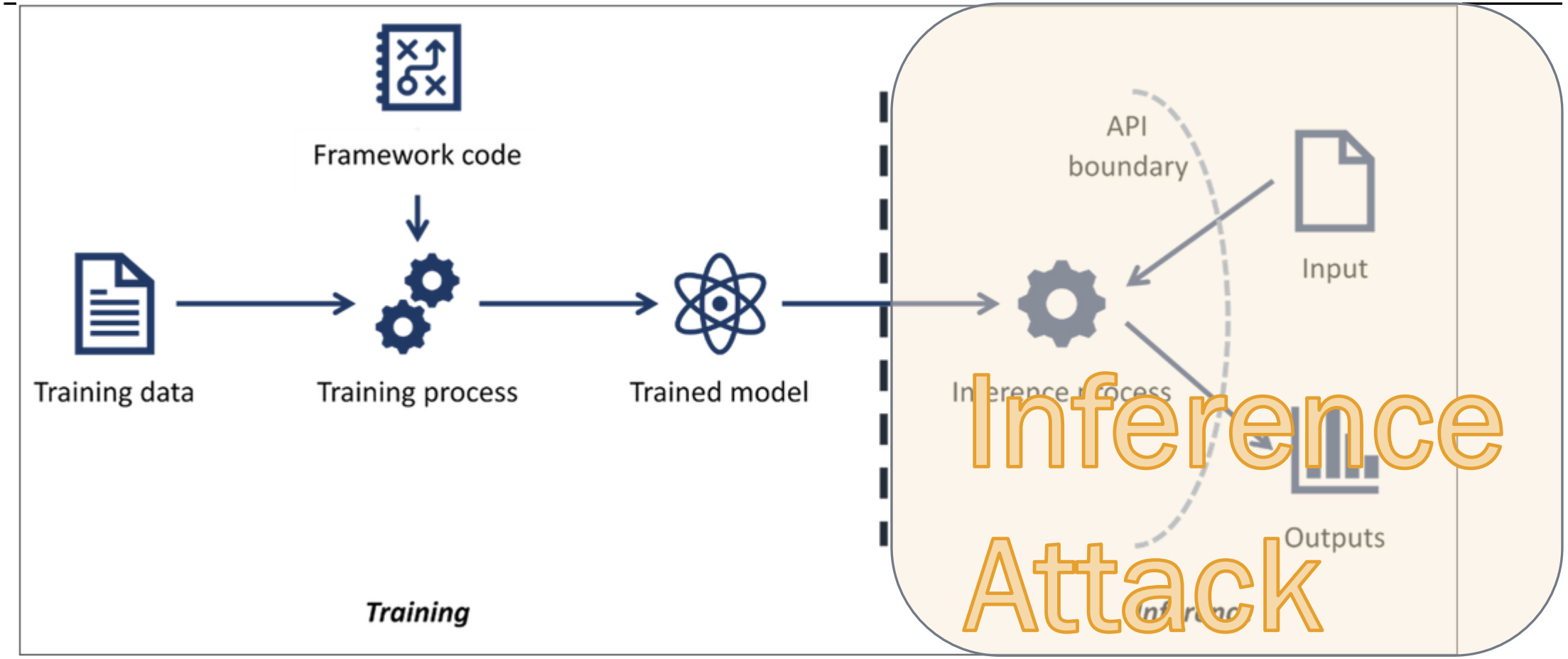


Image: https://github.com/mitre/advmlthreatmatrix/blob/master/images/AdvML101.PNG

ML Attacks Types

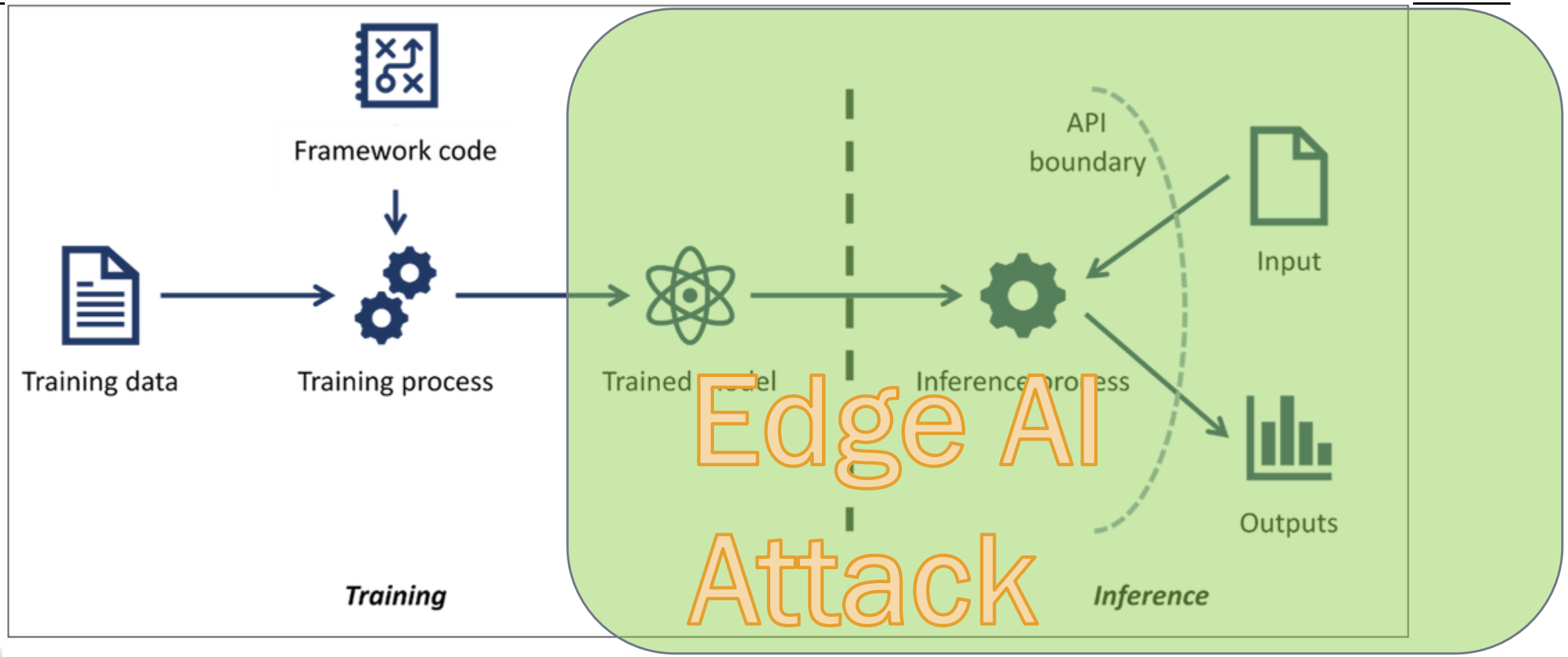


Image: https://github.com/mitre/advmlthreatmatrix/blob/master/images/AdvML101.PNG

Introduction to Cybersecurity & Cyber Conflict

History

Vulnerabilities

Attack/Defend Concept

Cyber Law

Cybersecurity Workforce

Landscape:
- Commercial
- Government
- Academia

Threat Intelligence

Threat Hunting

Tools (Offensive / Defensive)

What's Next
- Cyber Careers
- Training resources
- Research areas

Next Course: Dec 1, 2020
EventBrite.com
Discount: GO-ISSA

tomaspena.53@[gmail.com]