AI vs AI

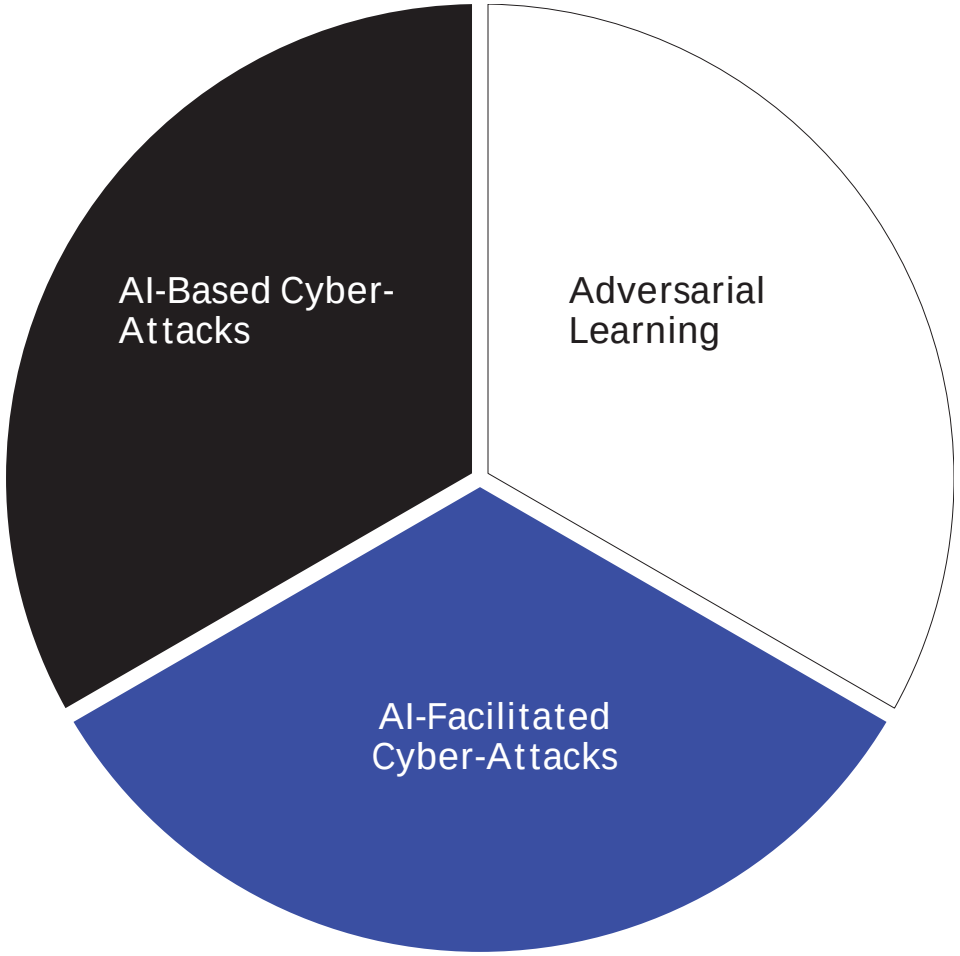
Think about what would happen when attackers start using the power of deep learning and machine learning to their advantage.

Nadav Maman, CTO | May 2020

deepinstinct™

What Does an AI-Based Attack Look Like?

AI Vs. AI: 3 Main Use Cases



- The malware operates AI algorithms as an integral part of its business logic.
- The malicious code and malware running on the victim's machine does not include AI algorithms, however AI is used elsewhere in the attacker's environment and infrastructure:
  - On the server side
  - In the malware creation process.
- The use of "malicious" AI algorithms to subvert the functionality of "benign" AI algorithms.

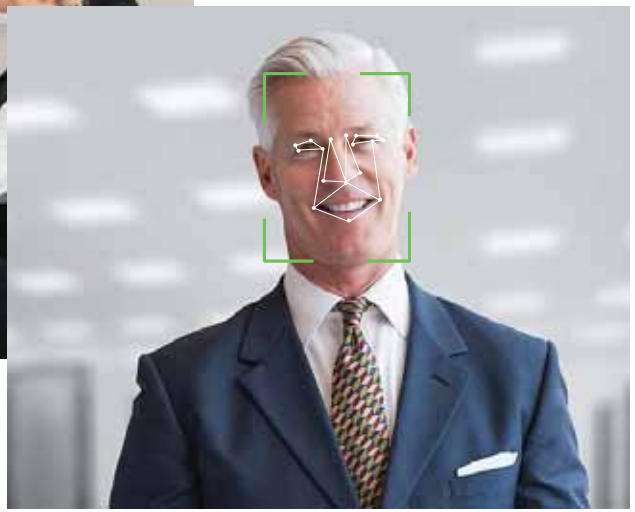
AI-based Cyber-attacks

- The malware operates AI algorithms as part of its business logic.
- In the past, such decisions could only be made manually by a human, as opposed to today, where it's able to be generated automatically.

AI-Based Cyber-Attacks

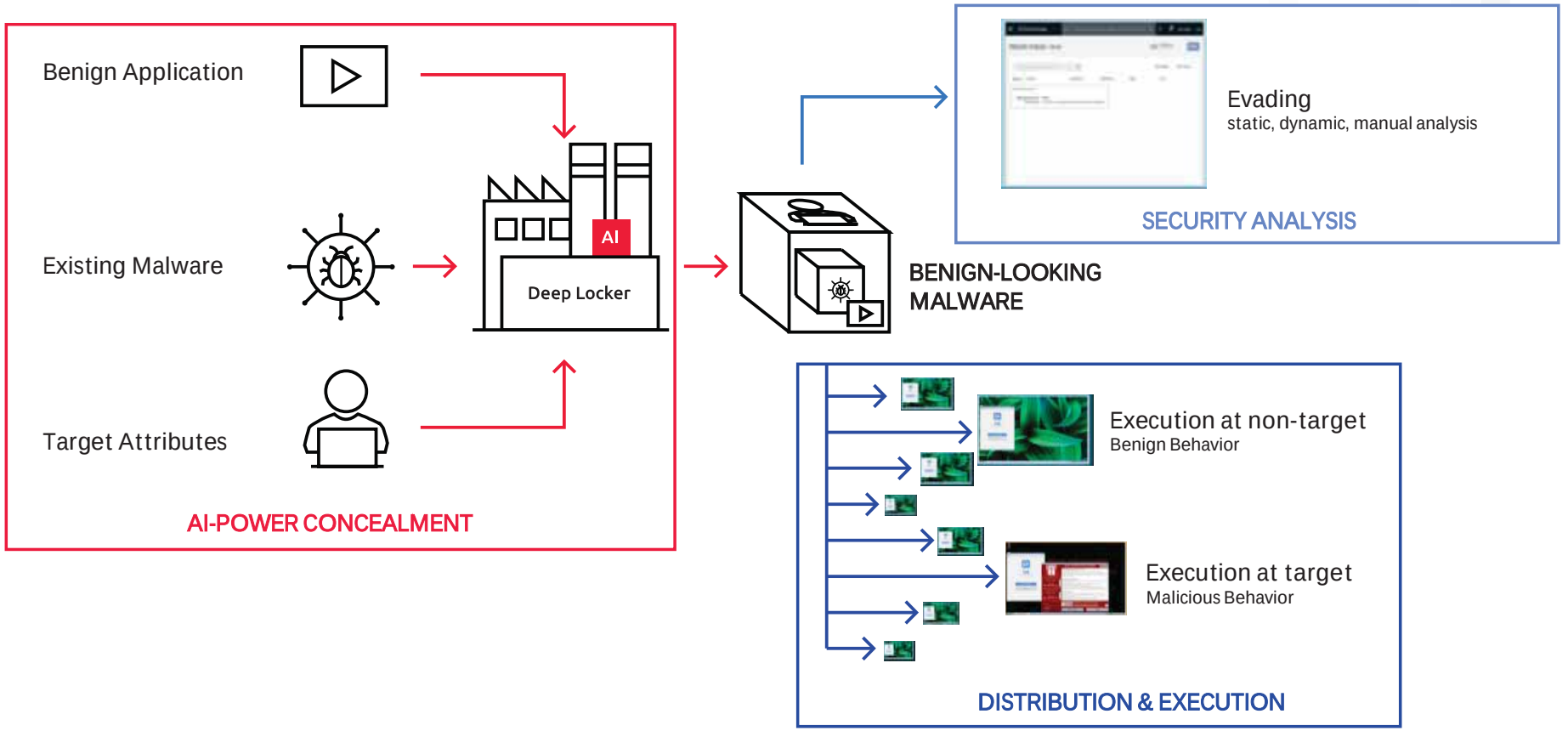
Example

- Deep Locker
  - An encrypted ransomware which autonomously decides which computer to attack based on a face recognition algorithm.

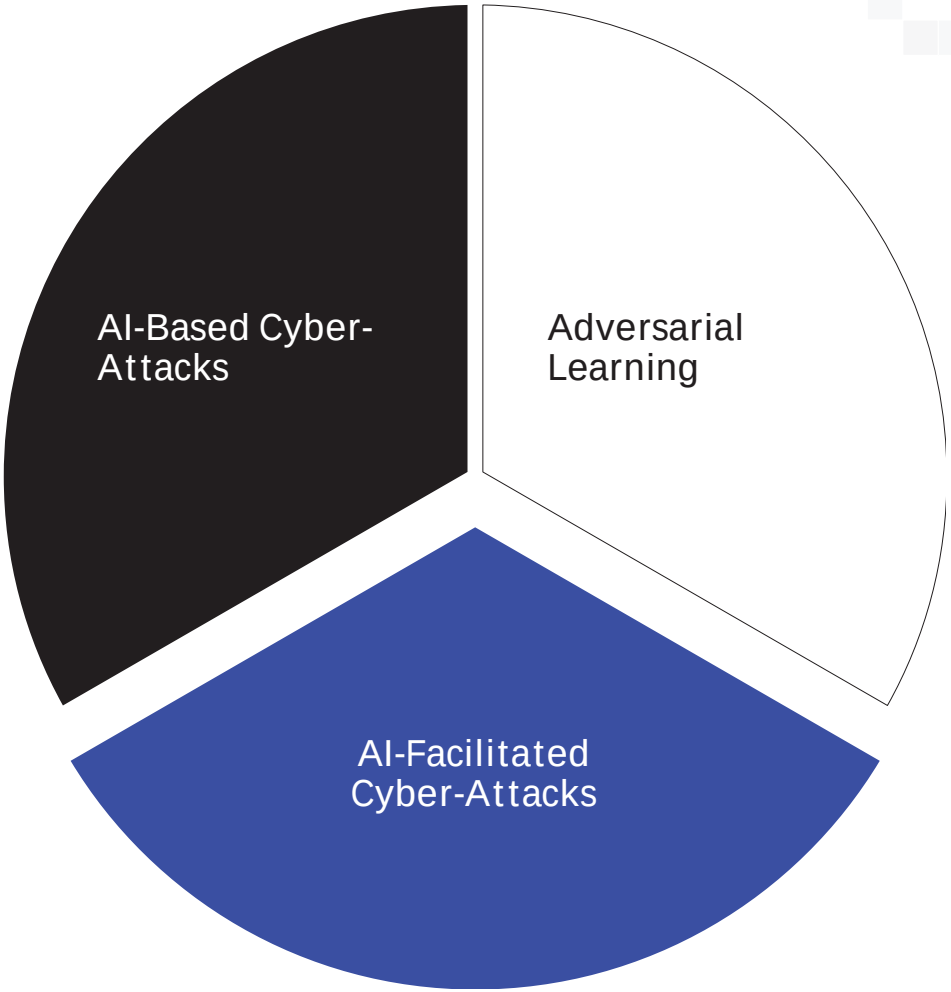


AI-Based Cyber-Attacks

Deep Locker



AI Vs. AI



AI Facilitated Cyber-attacks

- Malicious code and malware running on the victim's machine does not include AI algorithms, however AI is used elsewhere in the attacker's environment and infrastructure; be it on the server side, in the malware creation process etc.
- Infostealer – Sends endless data, which will be hard to sort based on human resources and classify it using AI.
- Images leaked from iPhone cloud – Assuming that you find a vulnerability and you would like to look for specific interesting images out of it, using image classification deep learning models.

2. AI Facilitated Cyber-attacks

Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter

High value targets on Twitter

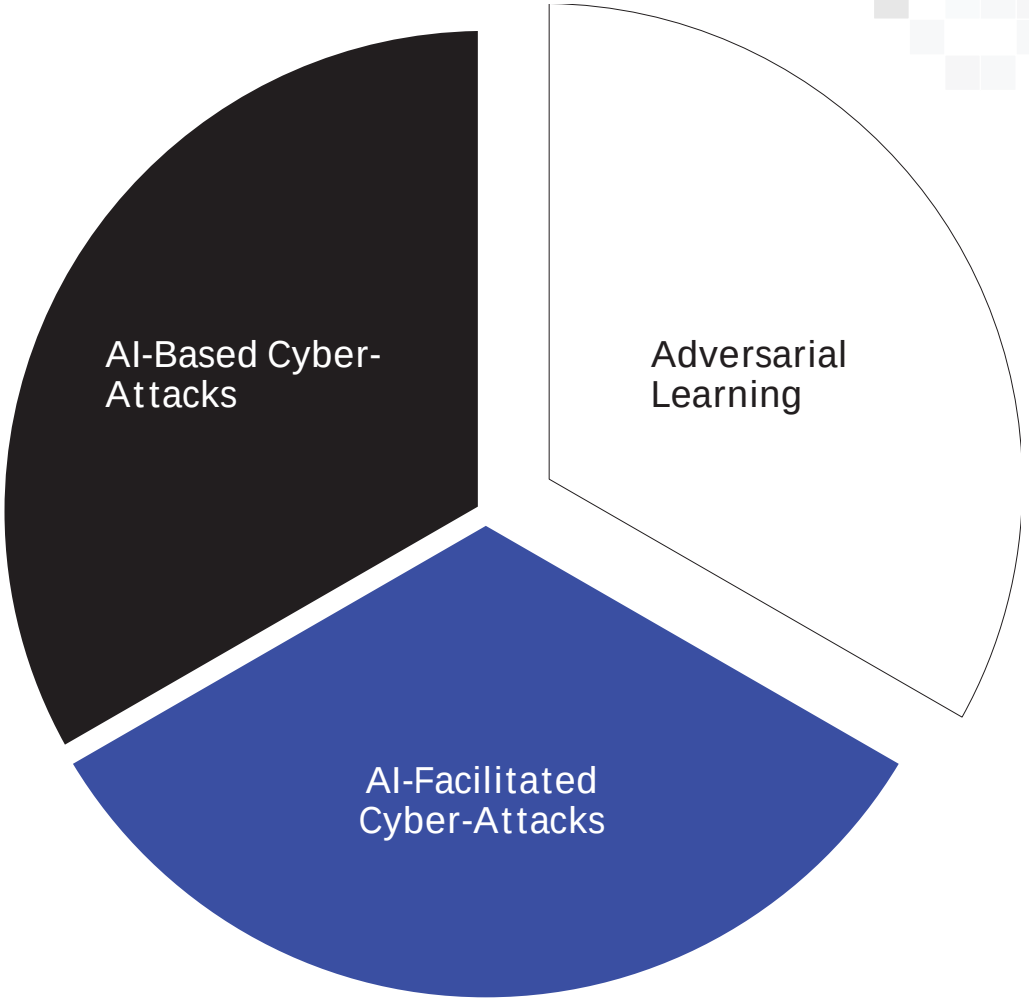
Feature extraction

Selects best clustering model

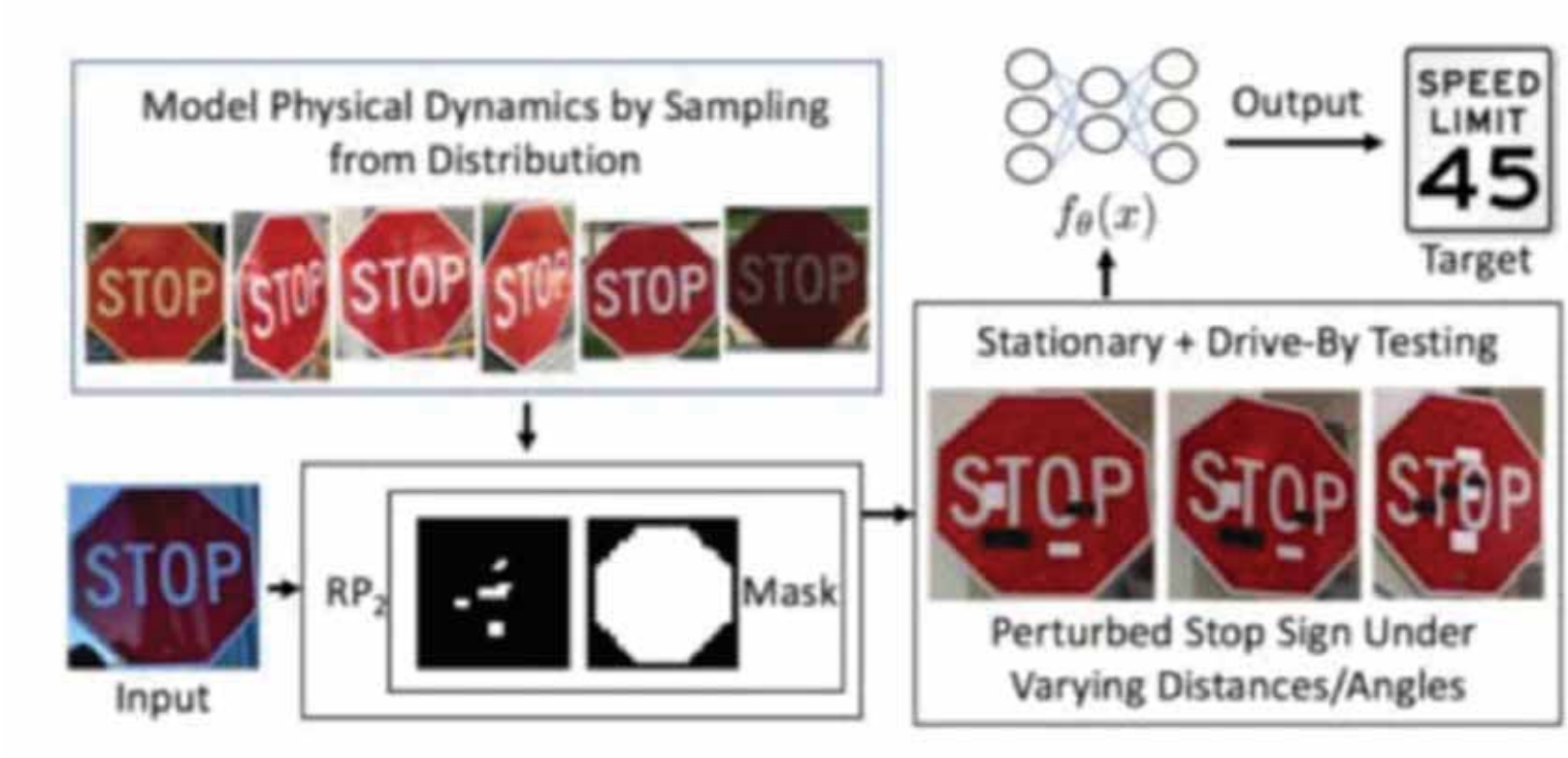
Automated spear phishing

Click rate measurement

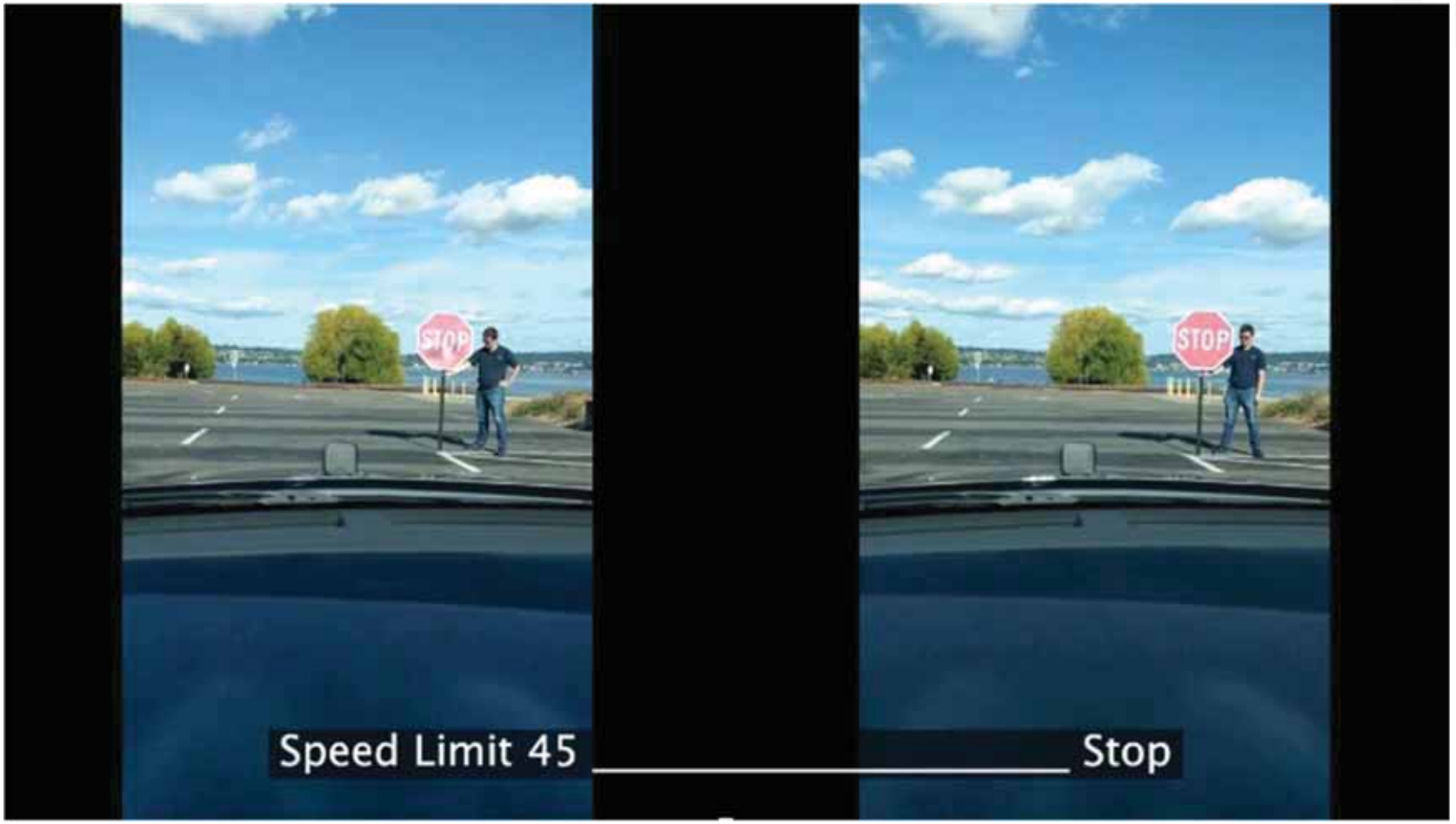
AI Vs. AI



3. Adversarial Attacks - Placing a Sticker in a Strategic Position on a Stop Sign



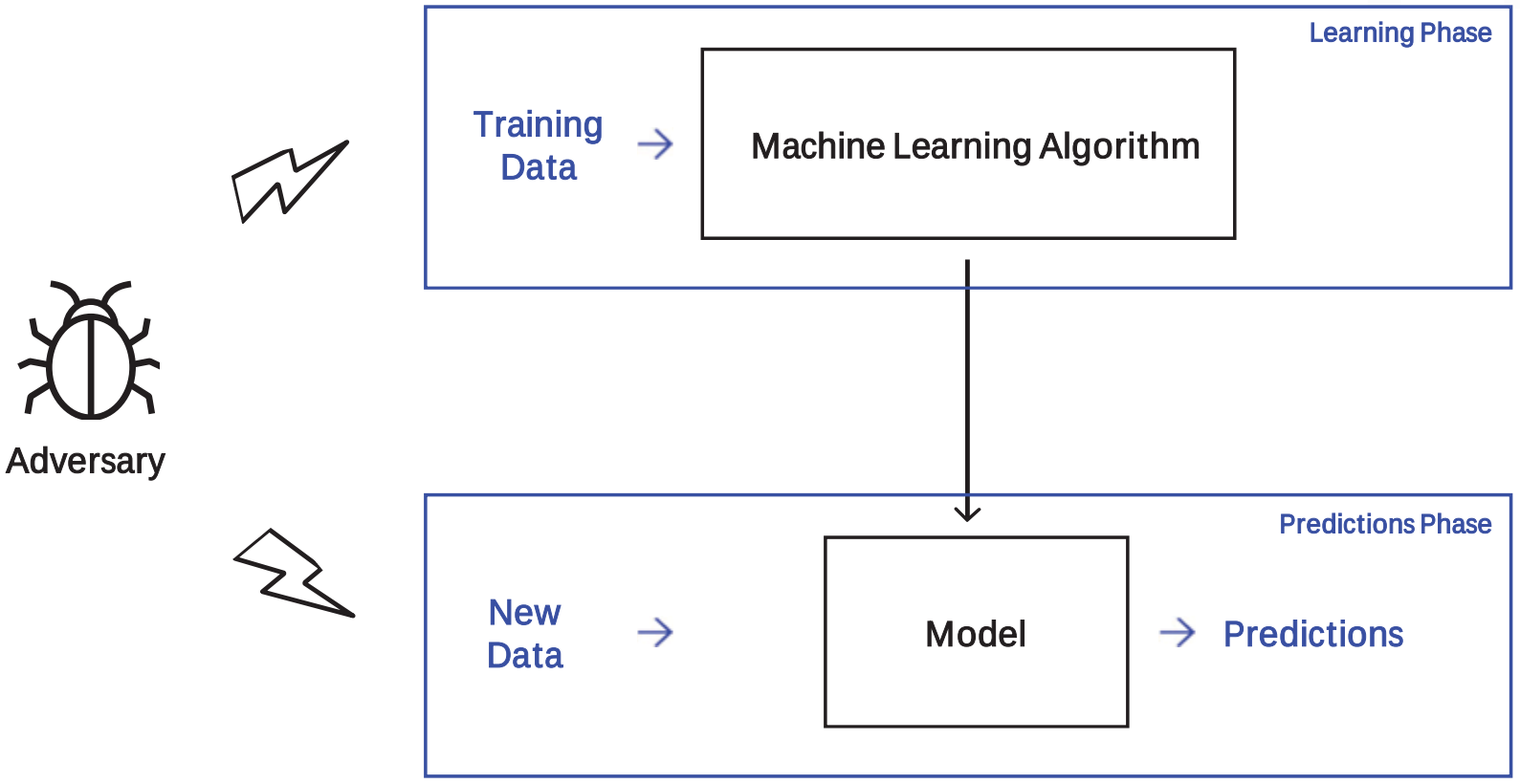
Adversarial Attacks



Which One is Which?

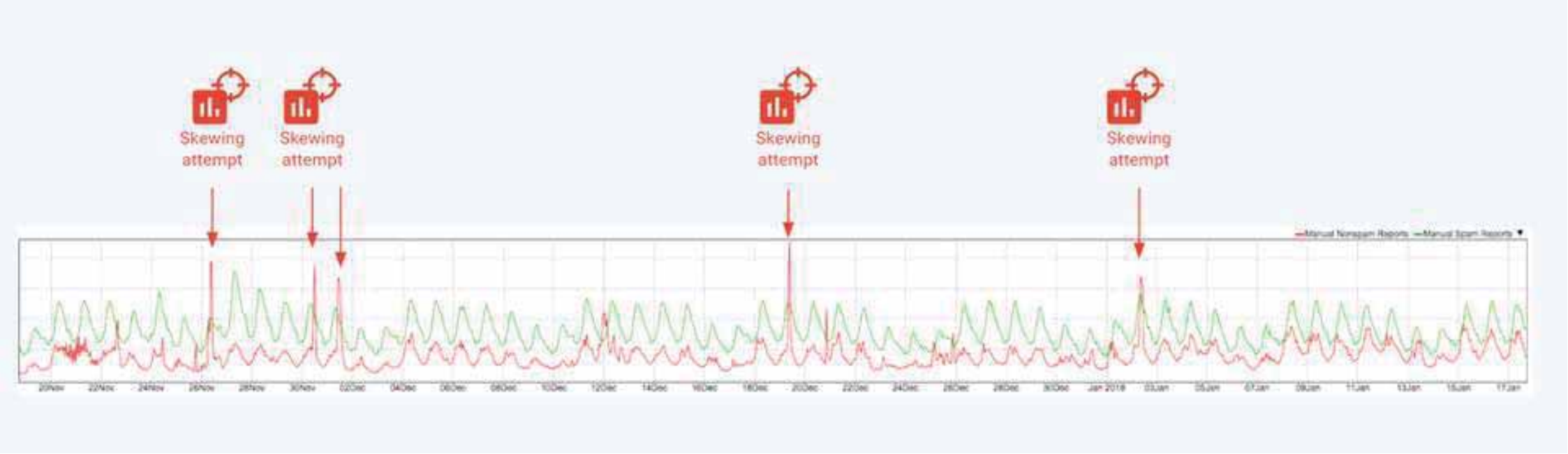
Adversarial Machine Learning

What options do attackers have to attack ML?



Poisoning Attacks

- For example, try to pollute training data to trick the classifier into marking specific malicious binaries as benign.



As shown in the figure, between the end of Nov 2017 and early 2018, there were at least four malicious large-scale attempts to skew Gmail filter off-track, by reporting massive amounts of spam emails as not spam.

Evasion Attacks

Attacker can try to:

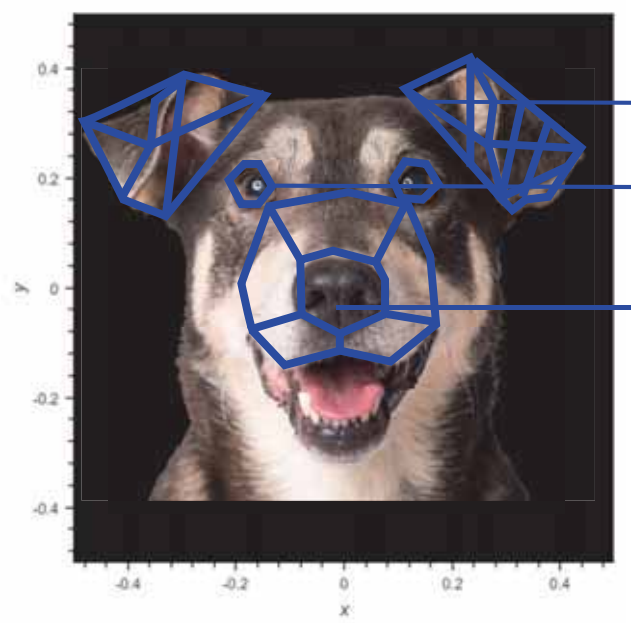
1. The attacker is trying to get a persistence on the machine

2. Retraining with adversarial examples, or “adversarial training” (RAD), by manipulating and changing the samples

16

Machine Learning – Feature Extraction

Dogs vs Cats



Ear = 9cm

Eyes = 4.2cm

Nose = 11.42cm

Benign vs Malicious

Counting

Floating points

Normalization

Binary features

Heuristics

How Feature Extraction can be Manipulated

Counting	→	Create additional sections
Floating points	→	Change time-stamps, pad the file with additional data
Normalization	→	Add plain data
Binary features	→	Pack malicious functionalities, create certificate
Heuristics	→	Additional evasion techniques
...		...

Deep Instinct - Deep Learning Cybersecurity Platform

Any Threat

- **File based threats:** PE, PDF, Office, fonts, TIFF, RTF, SWF, Mach-O, Macro, APK, Shellcodes
- **File-less based threats**: Macro, Scripts, Code injection, Dual-use
- **Ransomware**
- **Exploits**
- **Spyware**

Multi-Layered Protection

Pre-execution

- Deep Static analysis
- D-Cloud

On-execution

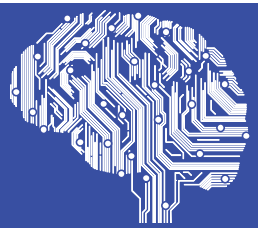
- Deep Behavioral analysis

Post execution

- Deep Classification
- Forensics & Remediation

Independent 3rd Party Tests

AV-TEST · AV Comparatives · SE Labs



ZERO-TIME THREAT PREVENTION PLATFROM POWERED BY DEEP LEARNING

PREVENT ▪ CLASSIFY ▪ CONTAIN

Compliance & Regulation

HIPAA · PCI Security Standards Council · amtso · GDPR

Anywhere

Cross-OS

Windows · macOS · Android · iOS · Chrome OS

Anywhere

- Endpoint
- Mobile
- Server
- Networks

Any Environment

- Online / Offline
- VDI
- Cloud / On-Premises
- Multi-Tenancy

Technology Partnerships

splunk> · MICRO FOCUS · amazon web services · IBM Radar · Citrix Ready · vmware · Microsoft Azure · Google Partners

Certification

Microsoft CERTIFIED · SE Labs · Google Cloud Platform

Do You Really Have a Defense Strategy in Place?

AI-based cyber-attacks

In the past such decisions could only be made manually by a human, as opposed to today, where it's able to generate decisions automatically.

Adversarial attacks against the usage of AI are **possible,** but not **feasible**

The attacker should know which features to use and know the model

Reverse engineer the model or use a predictive engine.

Much easier to attack predictable, high level features such as those used by our competitors.

Nearly all published adversarial attacks are for image recognition, not for cyber

Different size of files, etc.

You can't simply modify a raw byte (like you change a pixel's color) and expect the code to work.

We have defense methods implemented in our framework

Part of our IP

However, some of the research is available online.

THANK YOU

Nadav@DeepInstinct.com
CTO