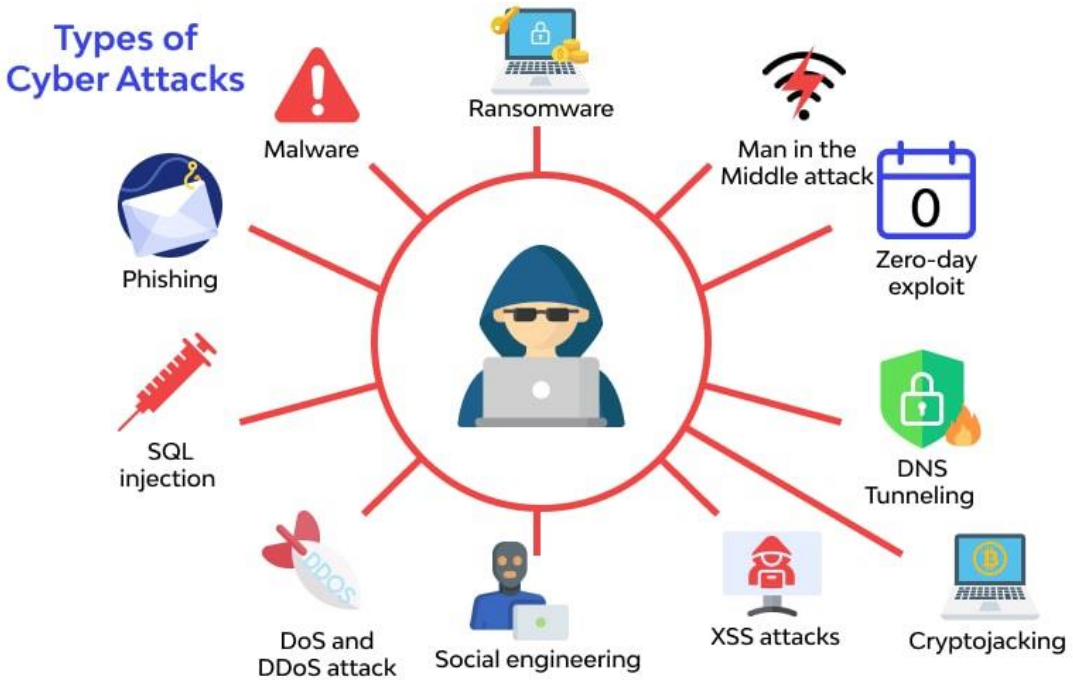
Why Machine Learning is needed for Cyber Security Research and Applications ?

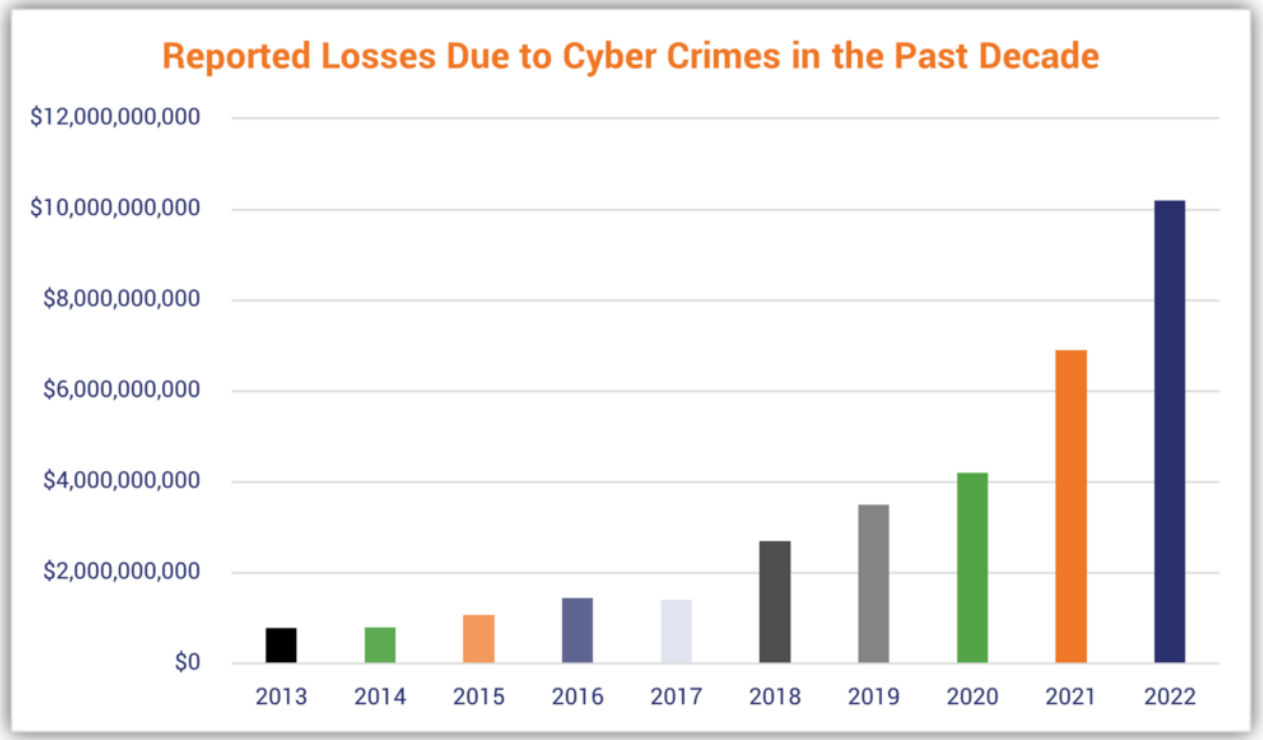
Yufei Han

CIDRE Team (https://team.inria.fr/cidre/)

November 7 @ Cyber Campus

Cyber Attacks on the rise of 2023



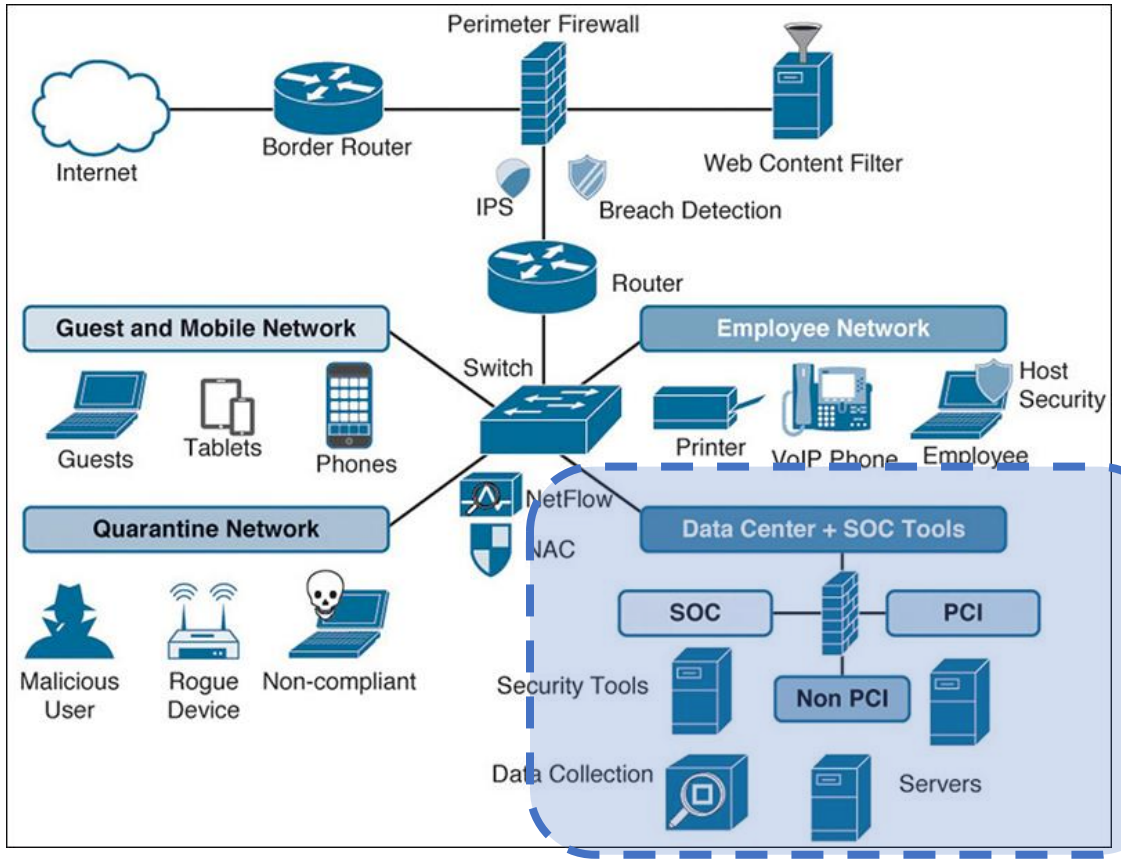


Cyber settacks causing increasingly expenseive loss across big to small, public to private entities

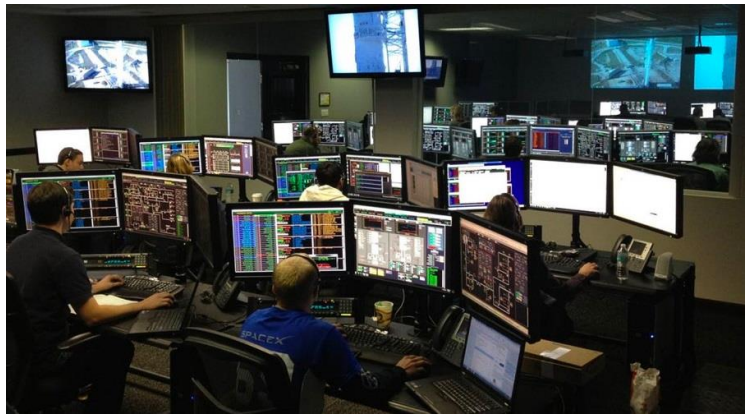
Why ML is needed for Cyber Security ?

Challenges to Cyber Security Service

Security **O**peration **C**enter (SOC) of Managed Security Service



Detection
Mitigation
Prevention
Monitoring



<u>Challenge raised by fast increasing cyber threats</u>:

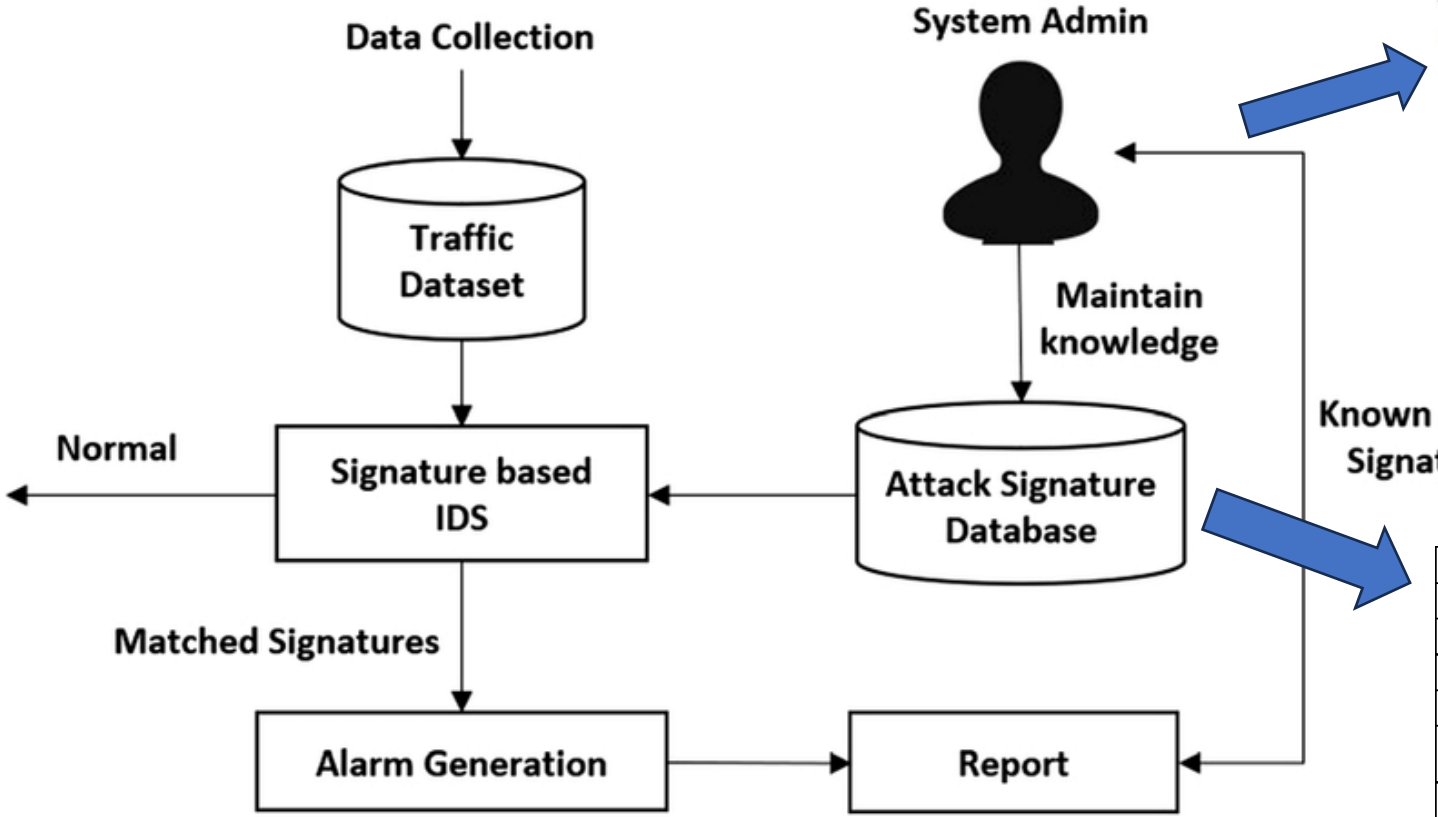
- **Huge volume of data input of SOC**. For example SOC of Norton recevie reports of 3.7 million spear-phishing and website hijacking events. **Human experts can not verify all of them.**
- 1/3 reported incidents originate from zero-day vulnerability. **Prediction of potential threats** is thus important for active defense
- **Privacy regulation: GDPR compliance in data usage**

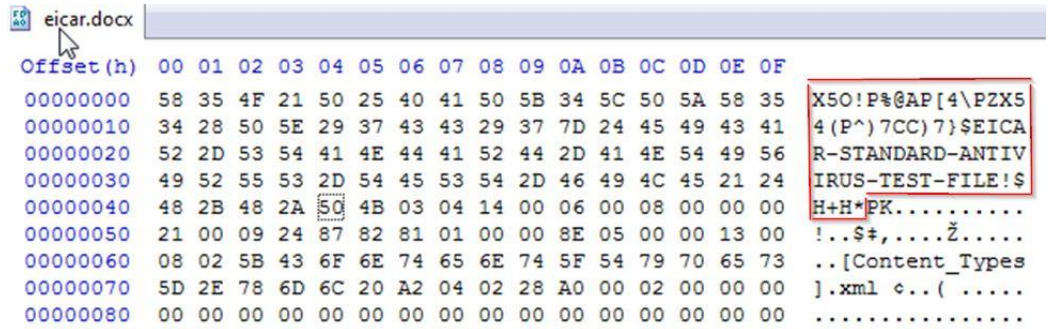
Why ML is needed for Cyber Security ?

Traditional approaches may fail

Signature-based cyber attack detection (like intrusion detection / malware classification)

Human experts explore signatures





Attack signature database

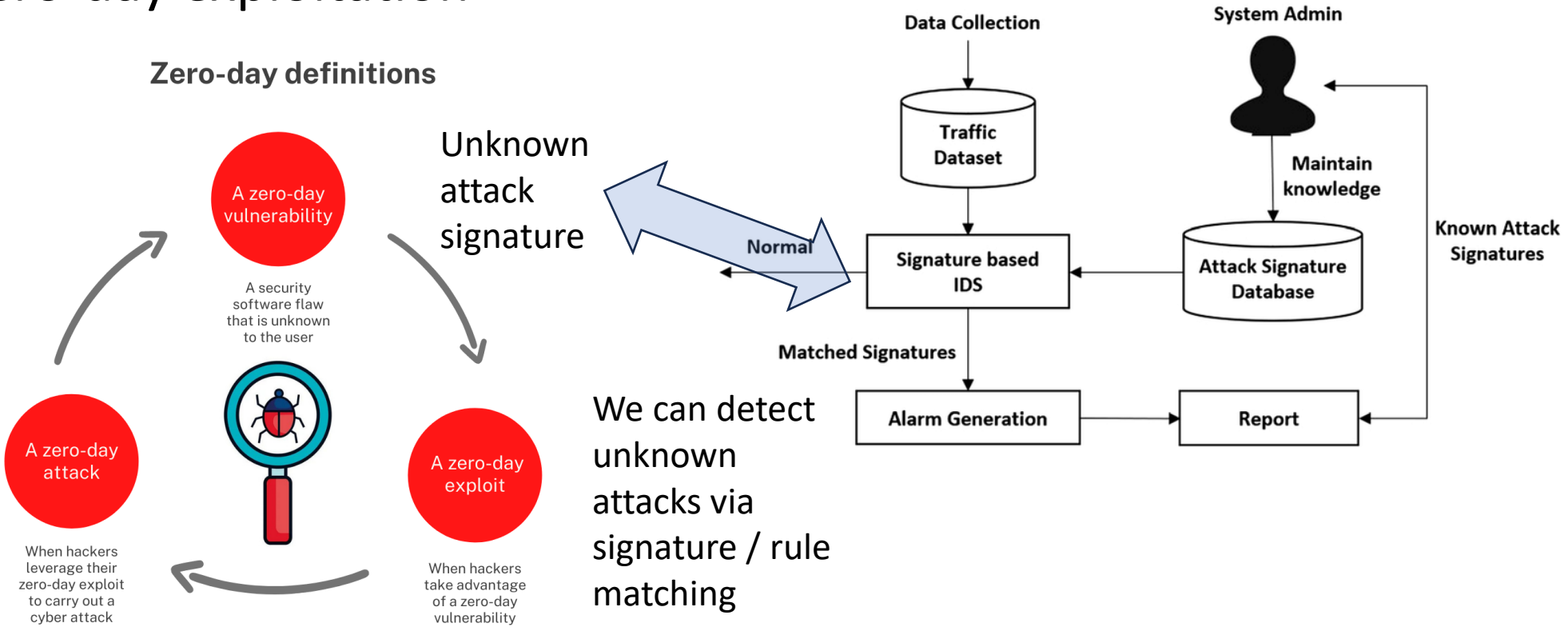
Virus name	Signature
Exploit.HTML.ObjectType	3c6f626a65637420747970653d222f2f2f2f2f2f2f2f2f2f2f
DOS.Arcv.1183	8dbc????b94c042e8005??47e2f9c3*33c0bb0001be0001899c????8984????e8
DOS.Trivial.Sbvc.A	e800005db824258d????????????23258d????cd21b44eb9
Worm.FlyStudio-20	5?5?5?5?f85?5?0f83
Worm.Allaple-315	c74424{3}(40|41)00{-10}8b4424{-30}33d2{-5}0401(5c|4c|44|54)24{-35}0401(5c|4c|44|54)24{-36}0401(5c|4c| 44|54)24
Dos.Flip.Gen	0ebb????????????b2??81c1????eb

Figure 1.Example virus signatures

Why ML is needed for Cyber Security ?

Traditional approaches may fail

- New attacks emerge rapidly: 1/3 of cyber attacks observed belong to zero-day exploitation



Why ML is needed for Cyber Security ?

Traditional approaches may fail

- High False Positive Rate: 25% to 75% alerts produced by SOCs are false positives
- Benign network traffics / executables could be flagged as malicious if it happened to share similar characteristics to a known threat.

Complex IT environments—SOC must monitor an ever-increasing number of alerts generated by all those security tools.

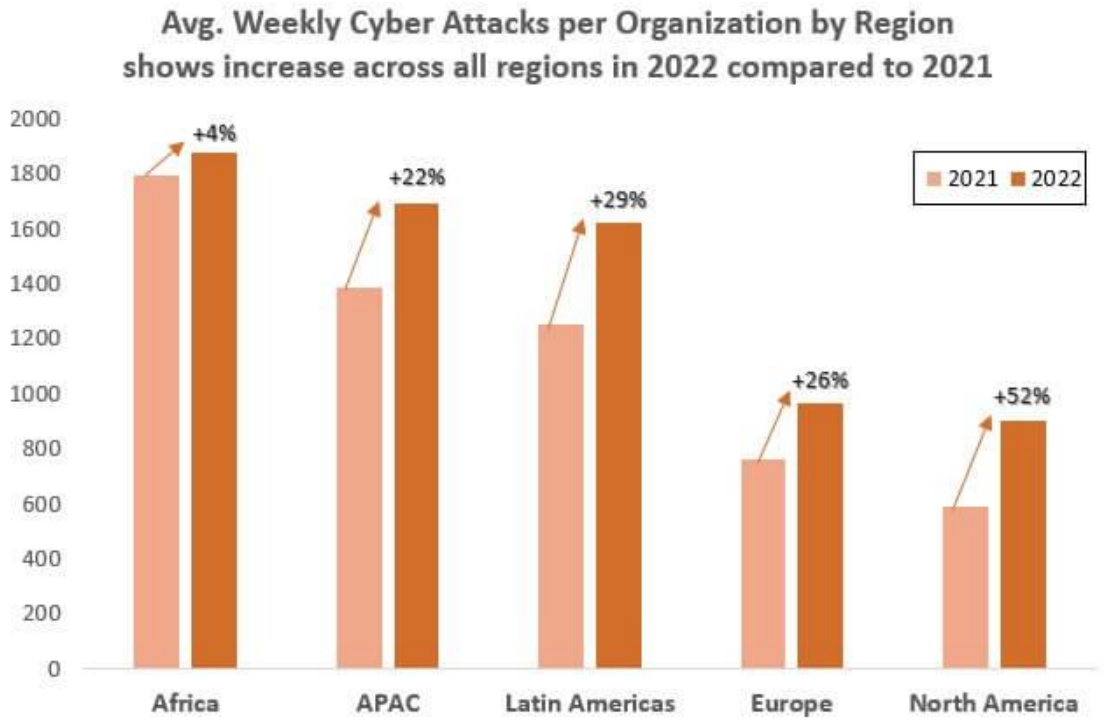
Lack of context—Alerts arrive, usually in someone's email inbox, without any context or actionable information.

Alert redundancy—Different security tools may be issuing alerts for the same issue.

Why ML is needed for Cyber Security ?

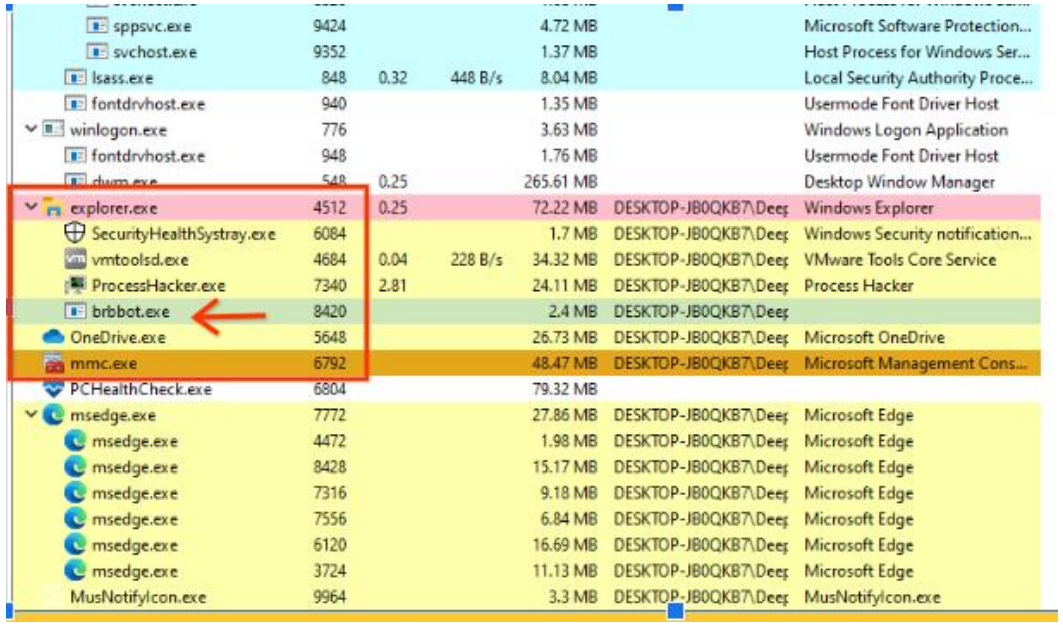
Traditional appraoches may fail

- Signature / rule-based detection heavilly rely on manual signature encoding

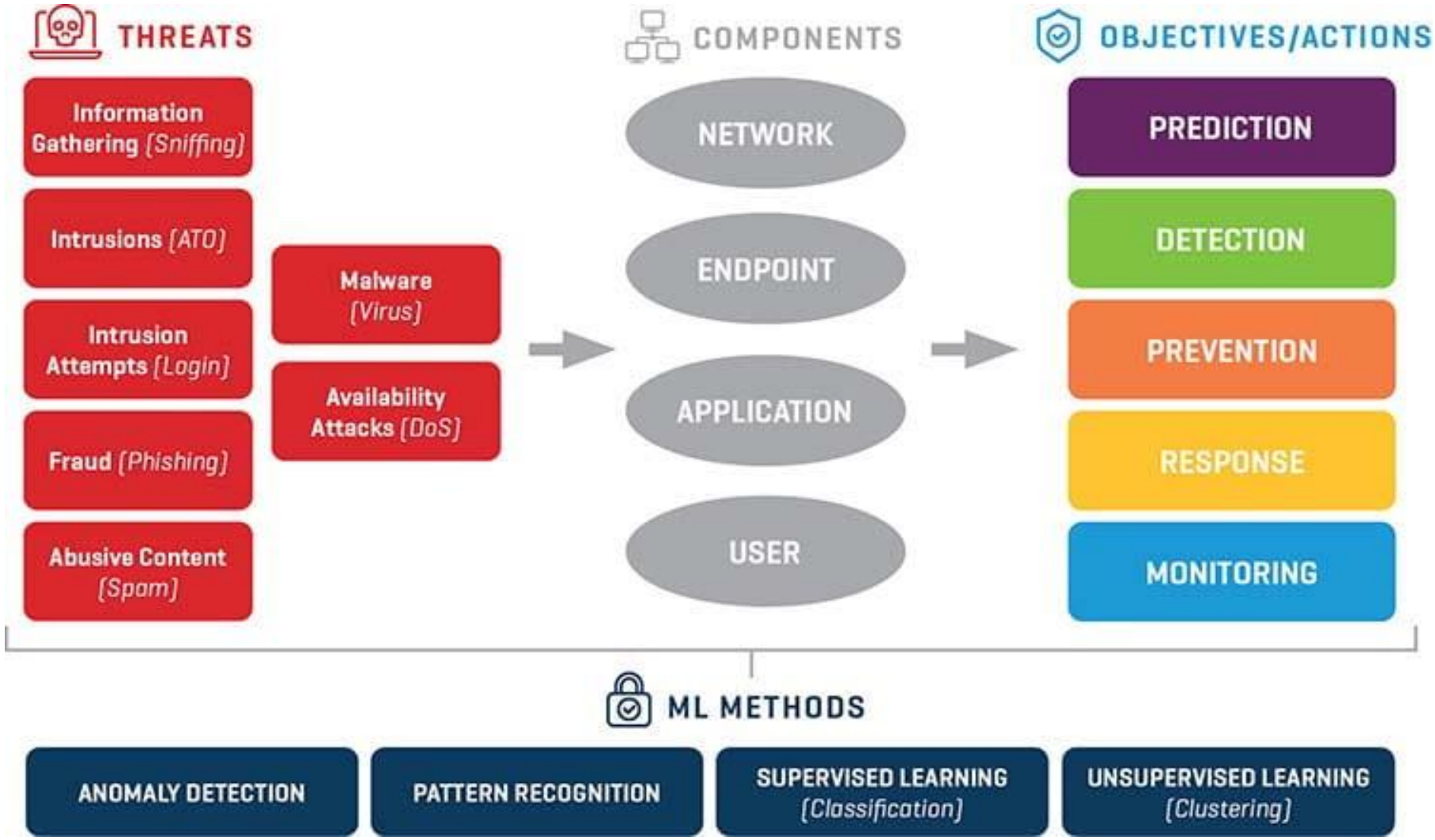


38% increase of 2022 global cyber attacks

SOC teams don't have enough human resources to address all of the reported incidents.



Use of Machine Learning in Cyber Security



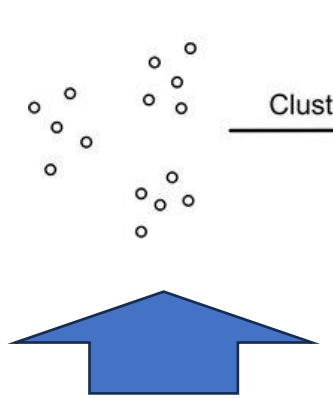
Why ML is needed for Cyber Security ?

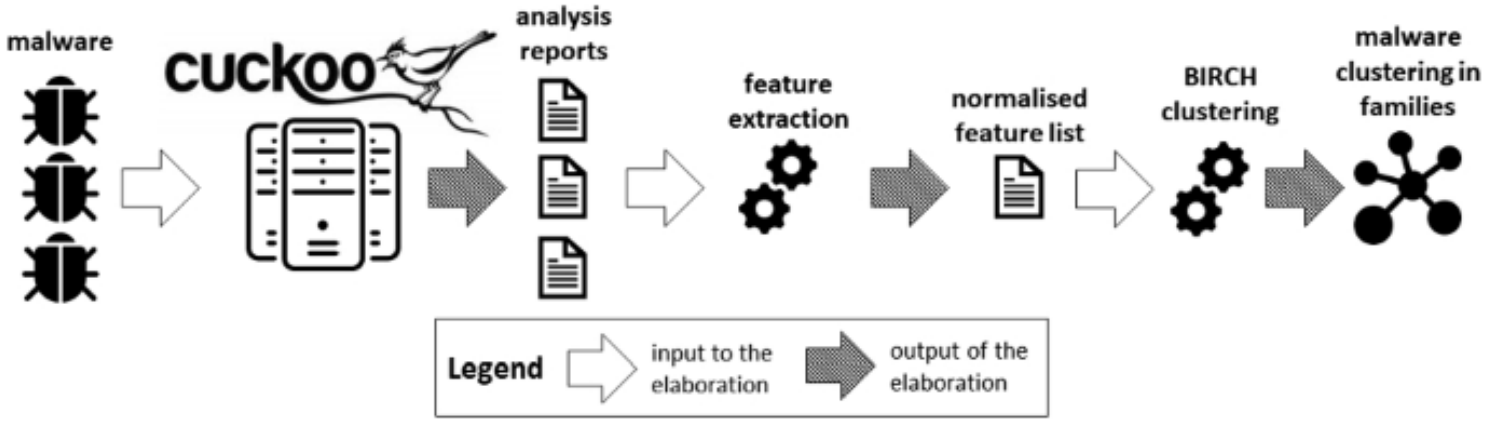
Use of Machine Learning in Cyber Security

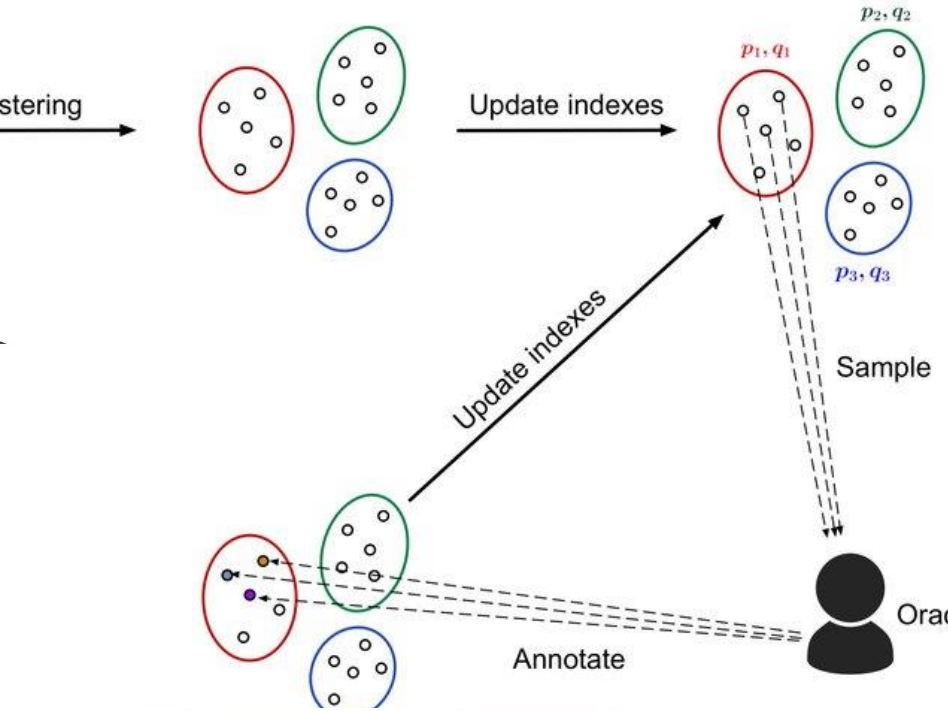
- Triage of attack campaigns

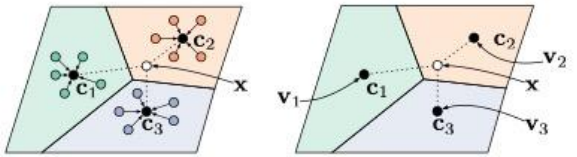
Self-Supervised Learning + Clustering: Triage attack campaigns in the observed security incident logs.

Active Learning: Reduce human manual investigation efforts (Human experts only need to focus on the most representative samples in each cluster)



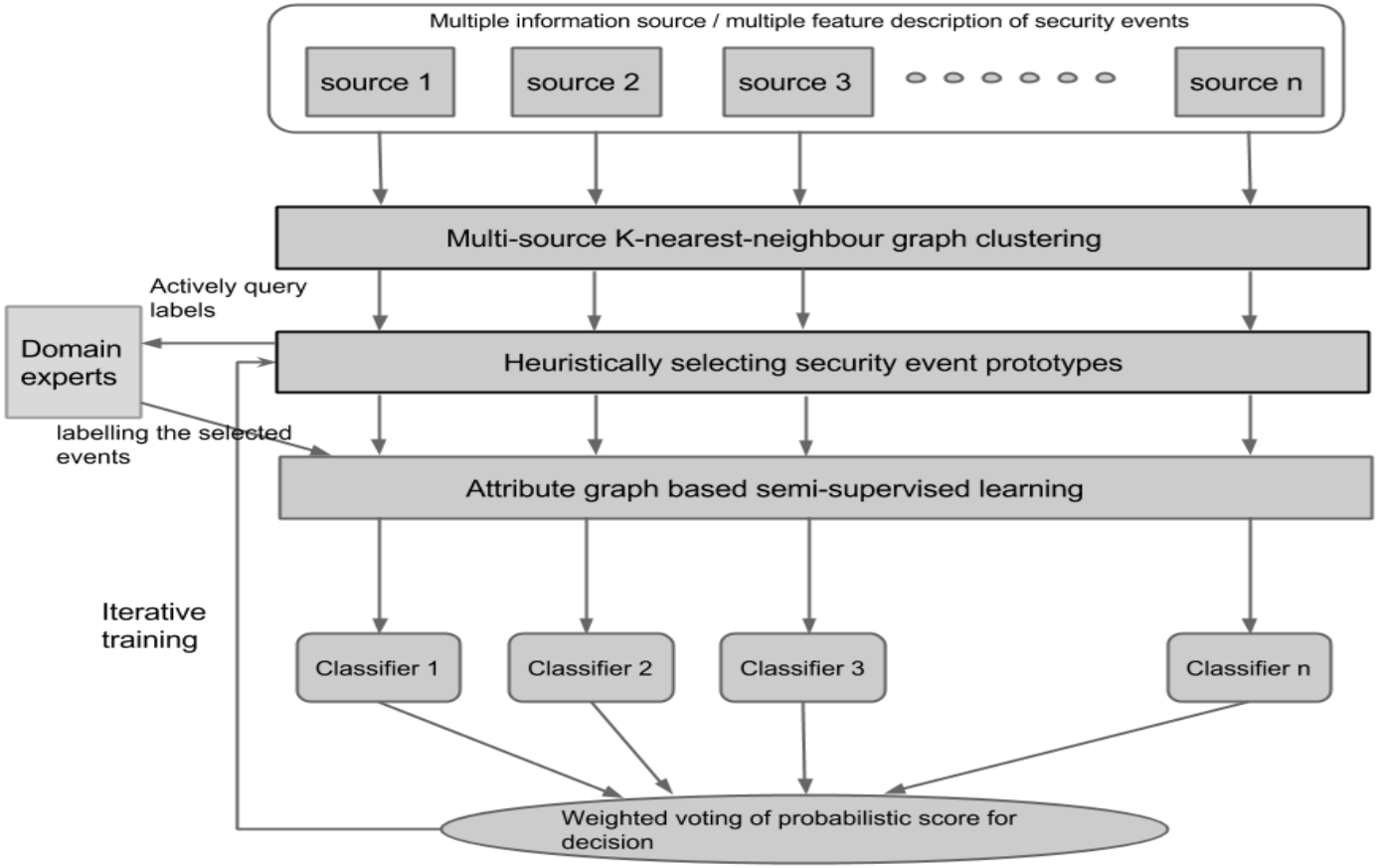






Use of Machine Learning in Cyber Security

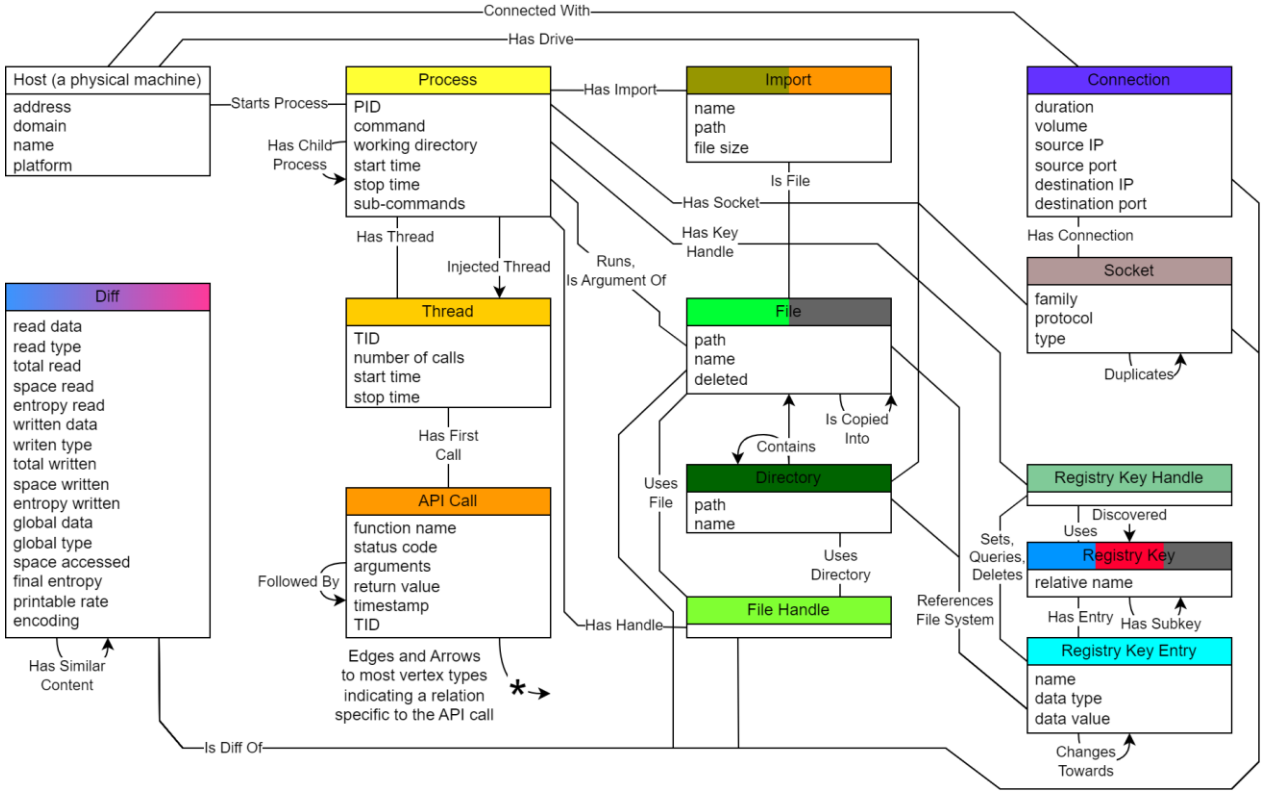
- Triage of attack campaigns



Real-world use case for phishing email classification: 660,000 suspicious emails, with only 0.01% of them labeled from 10 spear-phishing families, we are able to achieve 95% detection rate with FP = 0.08%, much better compared to rule-based detectors (Detection rate: 83% with the same FP)

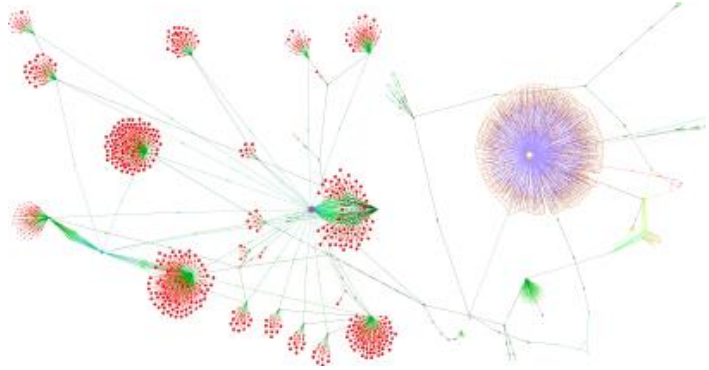
Use of Machine Learning in Cyber Security

- Unveil semantic explanation to cyber attacks



Structured representation of cyber attacks

Ransomeware Lockbit



Graph Neural Networks based subgraph matching

Unveil new attack behaviours

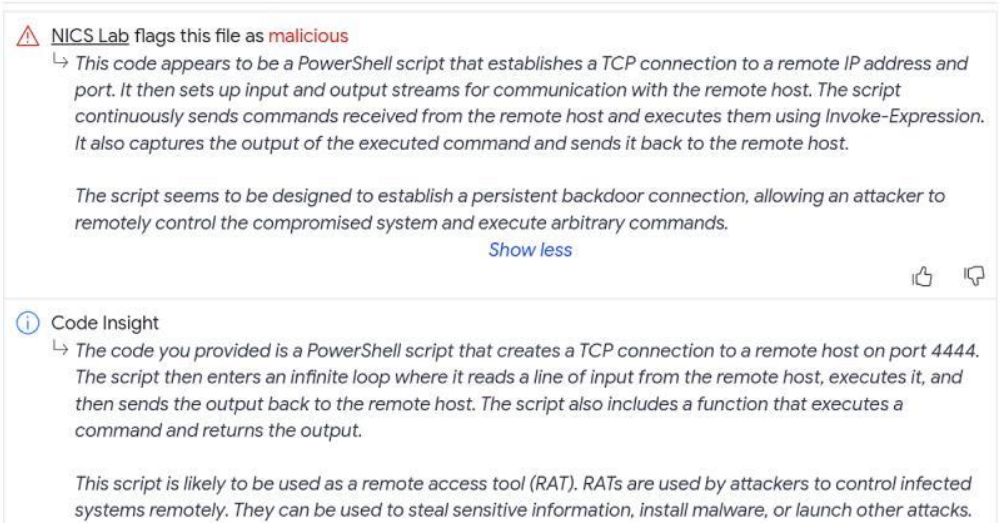
Detect evading techniques

Explaining attack process to reduce false alarms

Why ML is needed for Cyber Security ?

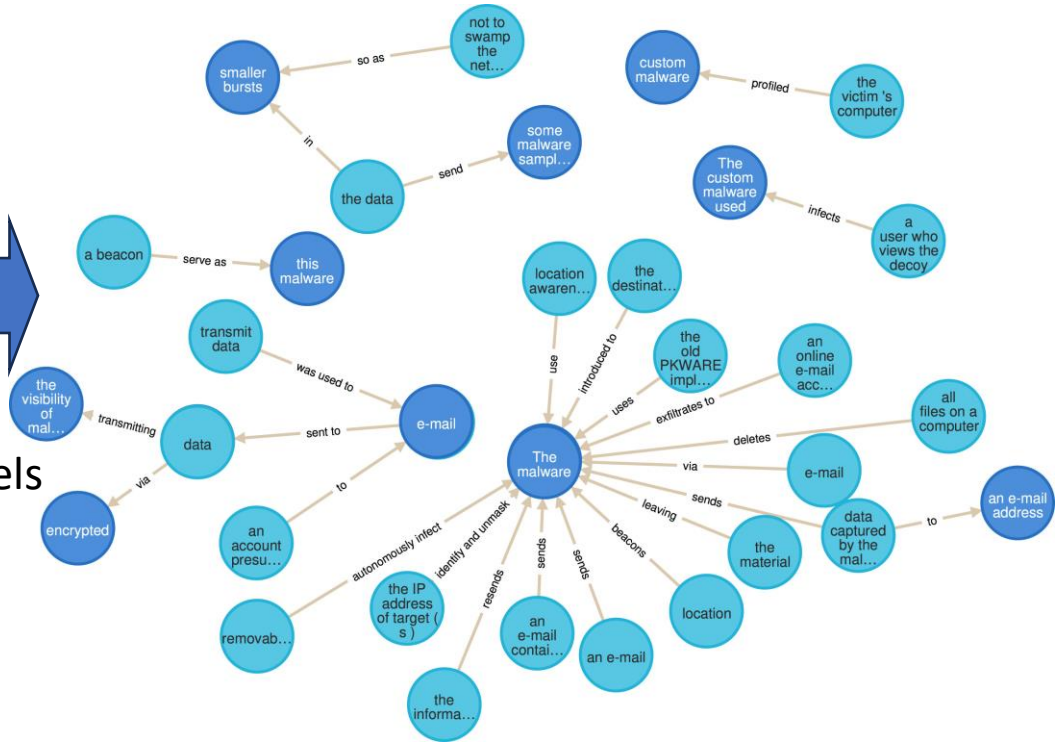
Use of Machine Learning in Cyber Security

- <u>Explore knowledge of cyber attacks and perform causality reasoning</u>



Threat intelligence reports of cyber attacks

Large Language Models



Knowledge Graph of cyber attacks

Why ML is needed for Cyber Security ?

Use of Machine Learning in Cyber Security

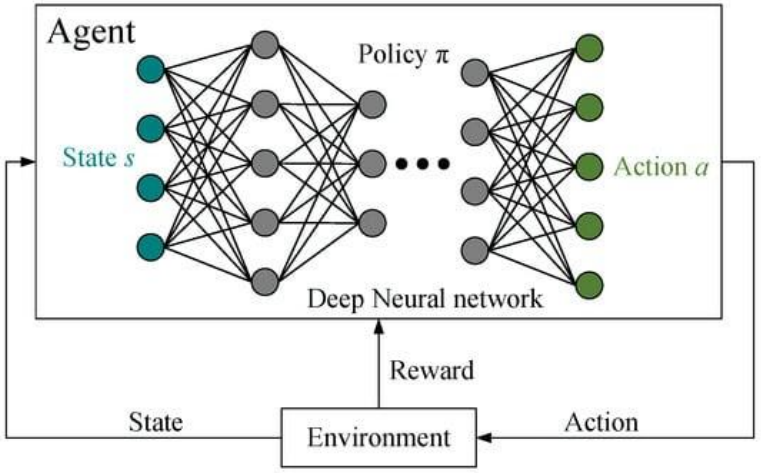
- Explore knowledge of cyber attacks and perform causality reasoning
  - Predict attack behaviours with causality reasoning
    - For remote access attacks, if we observe AnyDesk remote desktop application is installed onto the compromised device, we will be probably able to detect the execution of legitimate programs such as AmmyyAdmin and Team Viewer for remote interactive C2 to the victim device. Finally, the victim device will be controlled remotely via vncDll module or cloud service apps like LogMeIn. (Executable File Dropping – C2 connection – Remote Control session established)

    - Risk Assessment and Proactive Mitigation
      - Cyber Insurance
      - Actie Defense

Why ML is needed for Cyber Security ?

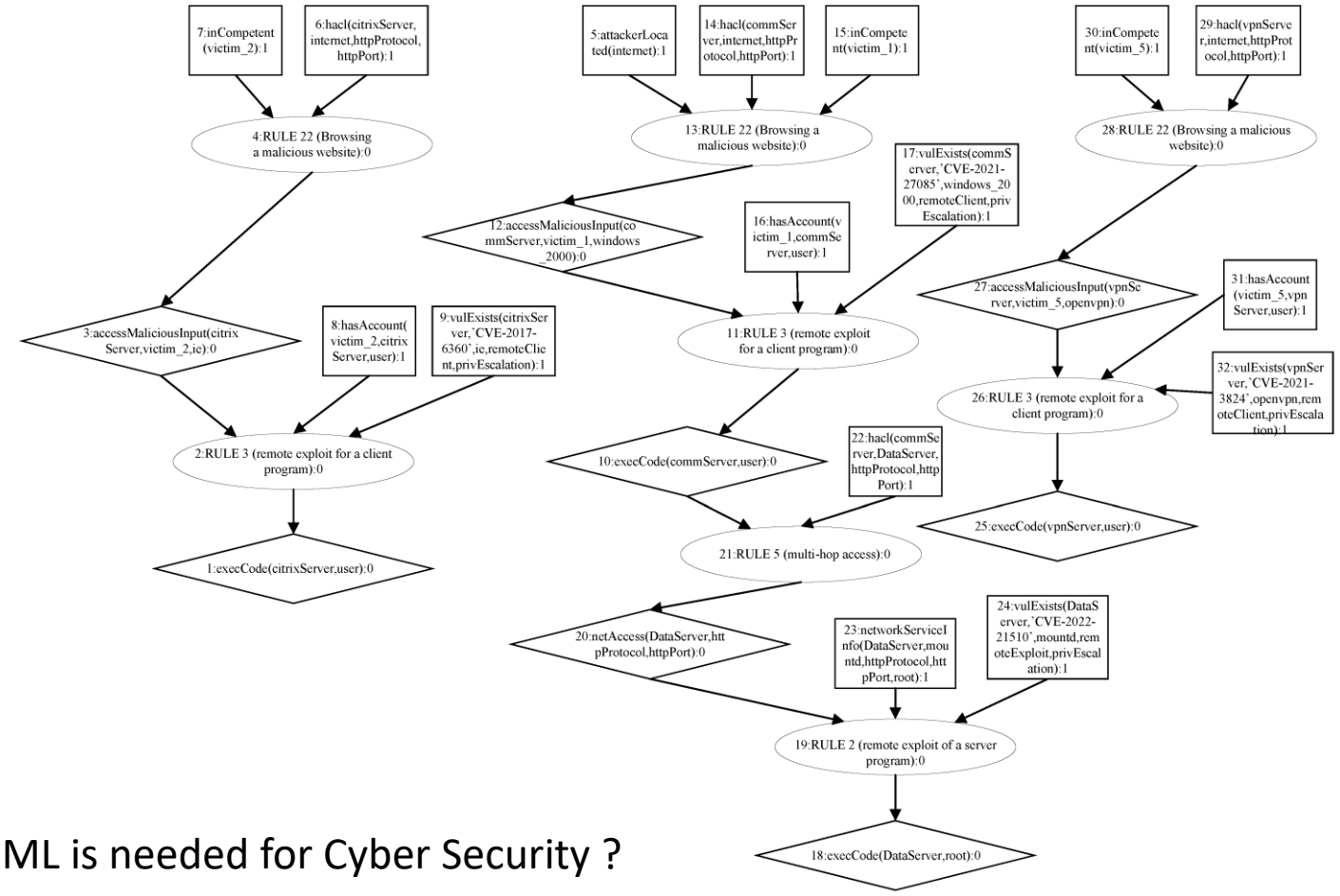
Use of Machine Learning in Cyber Security

- Automated Penetration Test



Deep Reinfrocement Learning as an automated agent to explore attack strategies

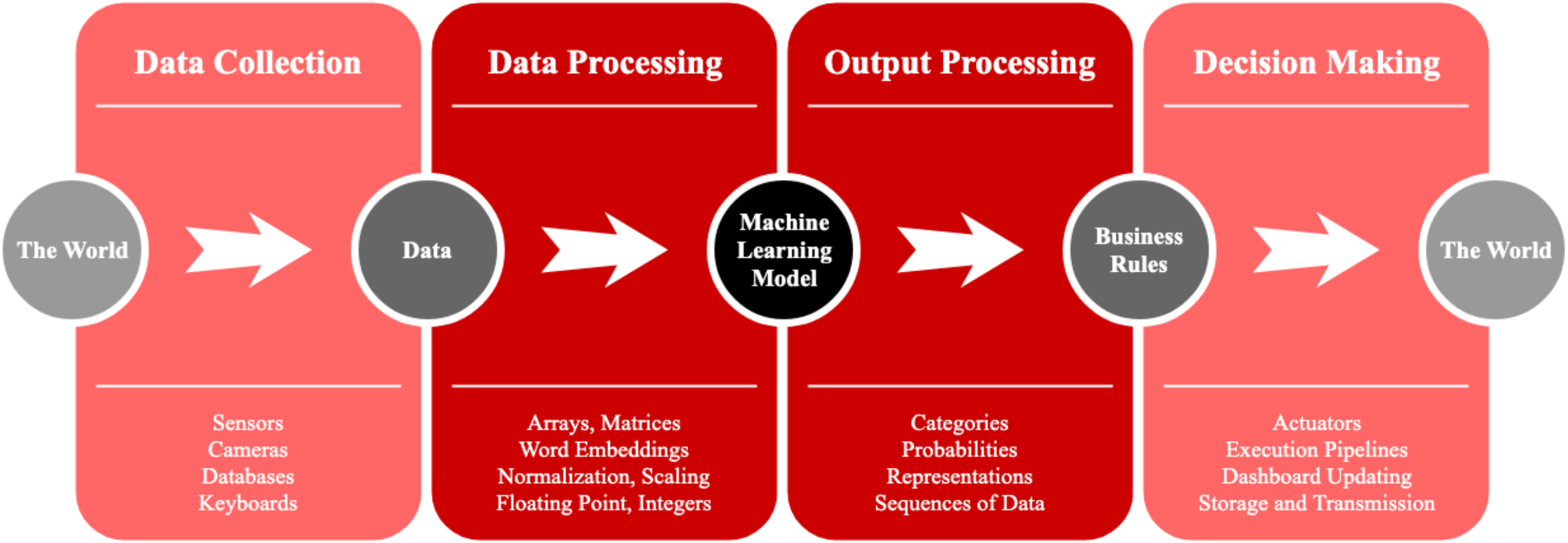
Attack paths unveiled by an automated pentest agent



Why ML is needed for Cyber Security ?

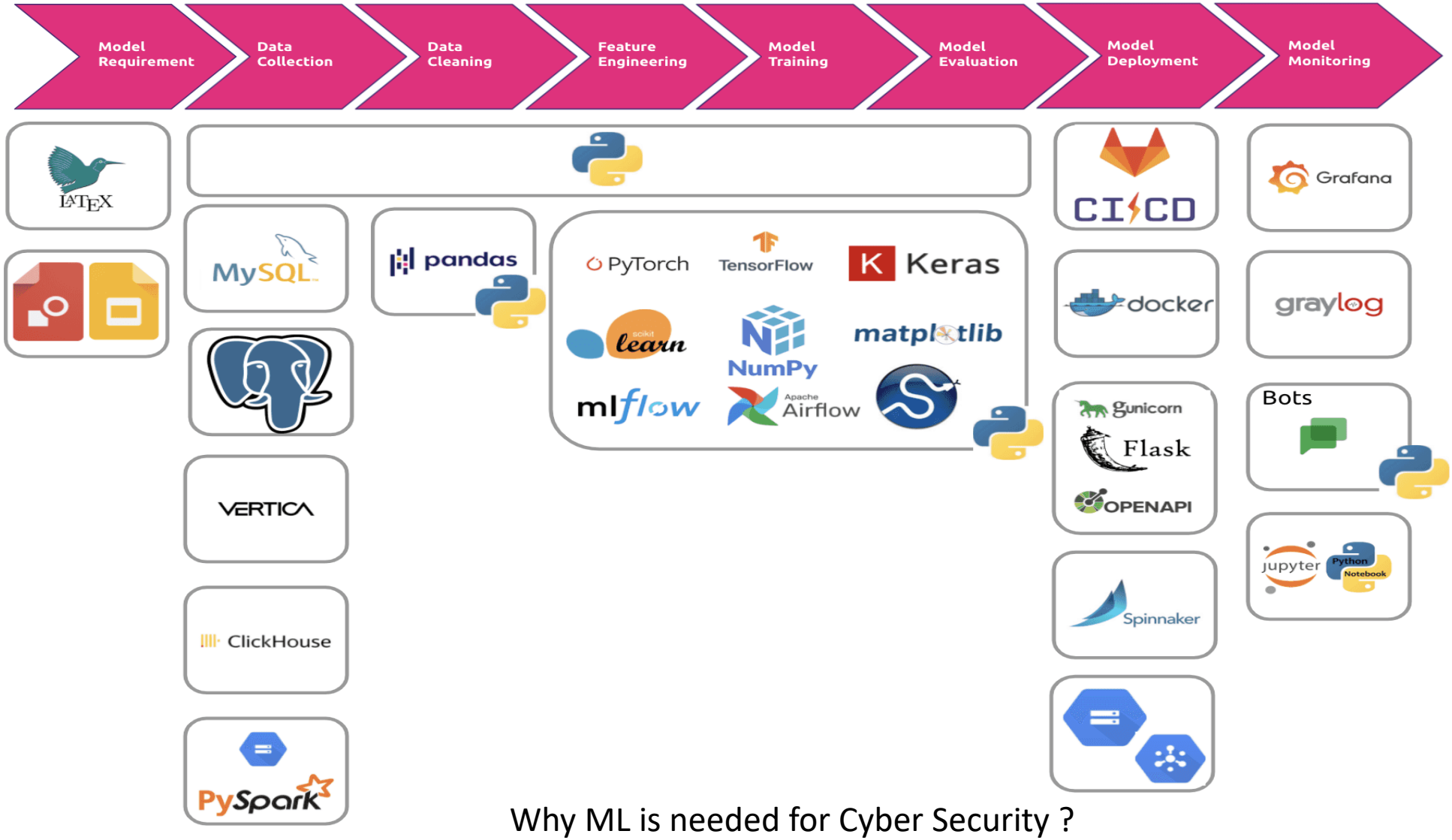
Machine Learning EcoSystem

A life cycle of a Machine Learning-based data analytical system



Why ML is needed for Cyber Security ?

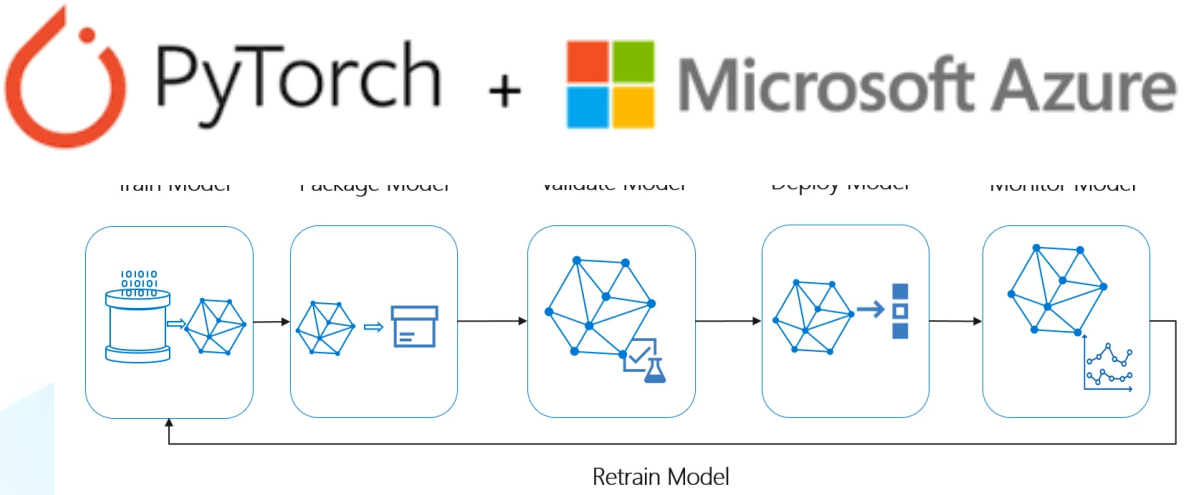
Machine Learning EcoSystem



Why ML is needed for Cyber Security ?

Machine Learning EcoSystem

- Concept proofing platform



Why ML is needed for Cyber Security ?

Thanks !

It is the Q&A time now ☺

https://team.inria.fr/cidre/