AI can Help Improve Network Security

From Better Attacks to Better defenses

Sebastian Garcia, Stratosphere Lab
AIC, CTU, CZ

AI *may* improve security

But before there are many questions:

- When?
- How?
- What do we need?
- What are we doing wrong?
- How good can we actually be at detecting?
- How good can we actually be at attacking?

When?

When? **Before** an incident

- New things
- Unkowns
- Attempts
- Reconnaissance
- Find vulnerable things
- Trends
- Impact measure
- Train

When? **During** an incident

- Was it completely successful?
- What is attacked?
- When did it started?
- From where? VPN?
- Who? Hacktivism? State?

- Is it contained?
- Got deep access?
- Miss something?
- Which technique?

When? **After** an incident

- Something missed
- Report for political/legal action
- TI gathering
- Prosecute
- Bigger fish

In which one are you now?

Before, during or after?

In all of them

AI Needs Network Security Datasets

Datasets are underestimated

- Research tends to be method-first.
- We do not usually evaluate if the data is good.
- We do not usually measure the bias in our data.
- We do not measure *what* we are missing.

Datasets. Benign

Getting malicious traffic is hard

Getting benign traffic is much harder

Datasets. Benign

- No clear definition of what it is
- Seasonality
- Cost of real labeling
- Privacy issues
- Legal issues
- Hard to publish. Anyone did?

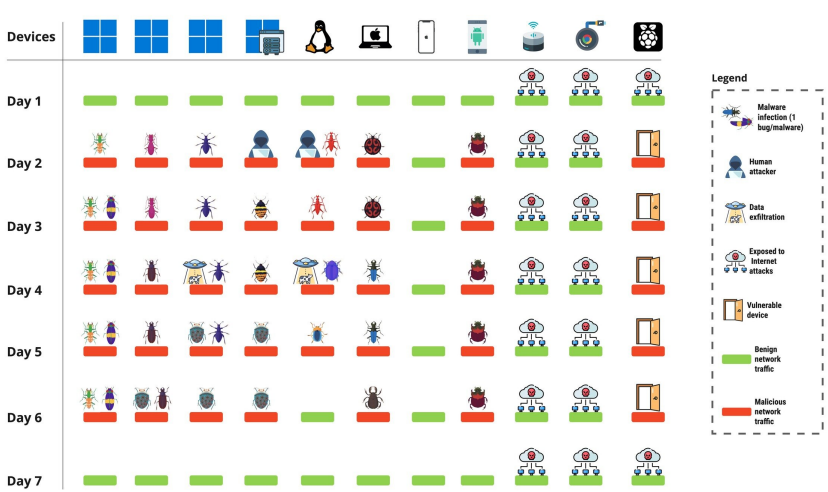
Datasets. Labels

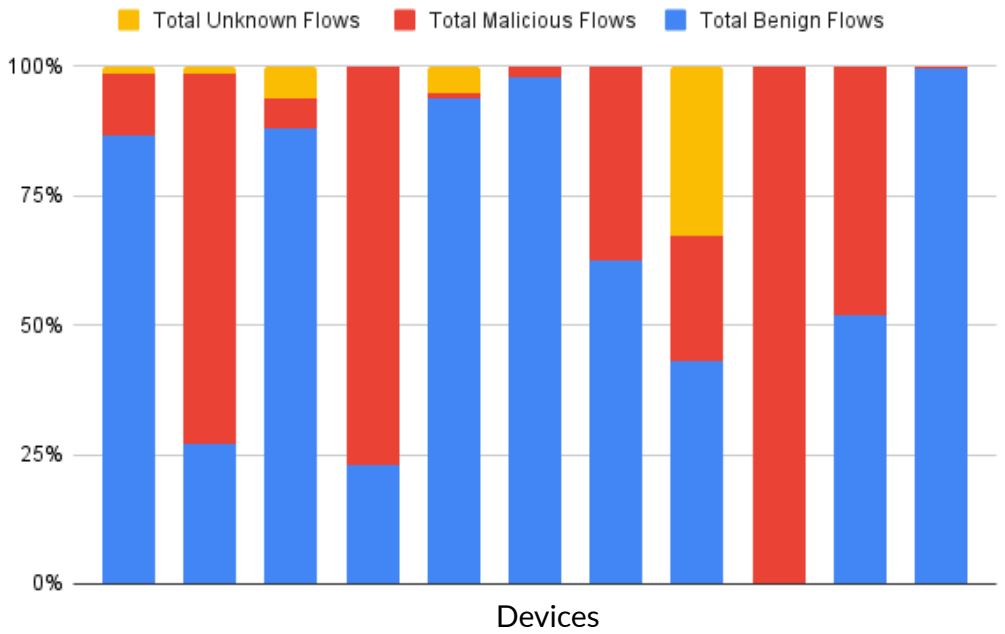
- The single most important commodity in datasets.
- Use experts for labeling.
- What are you labeling?
  - Src IP, dst IP, port, sequence, etc.
  - The same flow can have different labels
- Use tools, rules and ontology [1]

[1] https://github.com/stratosphereips/netflowlabeler

Datasets. Balance

- Bad ML requires 50/50 ratio of benign/malicious
- AD assumes >50% is benign





[1] CTU-SME-11 https://zenodo.org/record/7958259

Datasets are Not Enough

- Evaluate an attacker waiting?
- Evaluate a computer infected while being attacked?
- Evaluate IDS communicating between themselves?
- Evaluate the evolving TI feeds?
- Evaluate a human attacker taking decisions?

Detection with AI

Detection

We want to detect:

- All attacks
- All the time
- Without errors
- In real time
- And evolve
- And cheap
- Thank you

Detection

All attacks

Cohen, F. (1987). Computer viruses: Theory and experiments. *Computers & Security*, 6(1), 22–35. https://doi.org/10.1016/0167-4048(87)90122-2

4.1 Detection of Viruses

In order to determine that a given program '*P*' is a virus, it must be determined that *P* infects other programs. This is undecidable since *P* could invoke any proposed decision procedure '*D*' and infect other programs if and only if *D* determines that *P* is not a virus. We conclude that a program that precisely discerns a virus from any other program by examining its appearance is infeasible.

No, we can't probably do this one

Detection

All the time

- In the lifecycle of an attack/malware
- Different conditions

Yeah, we can probably do this one

Detection

Without errors

- As Cohen said, no perfect detection, so we will have errors.

No, we can't probably do this one

Detection

In real time

Yeah, we can probably do this one given enough hardware and money

Is Detection Hard?

Detecting some malicious is not hard

Detecting some malicious among **benign** is hard.

Detection depends...

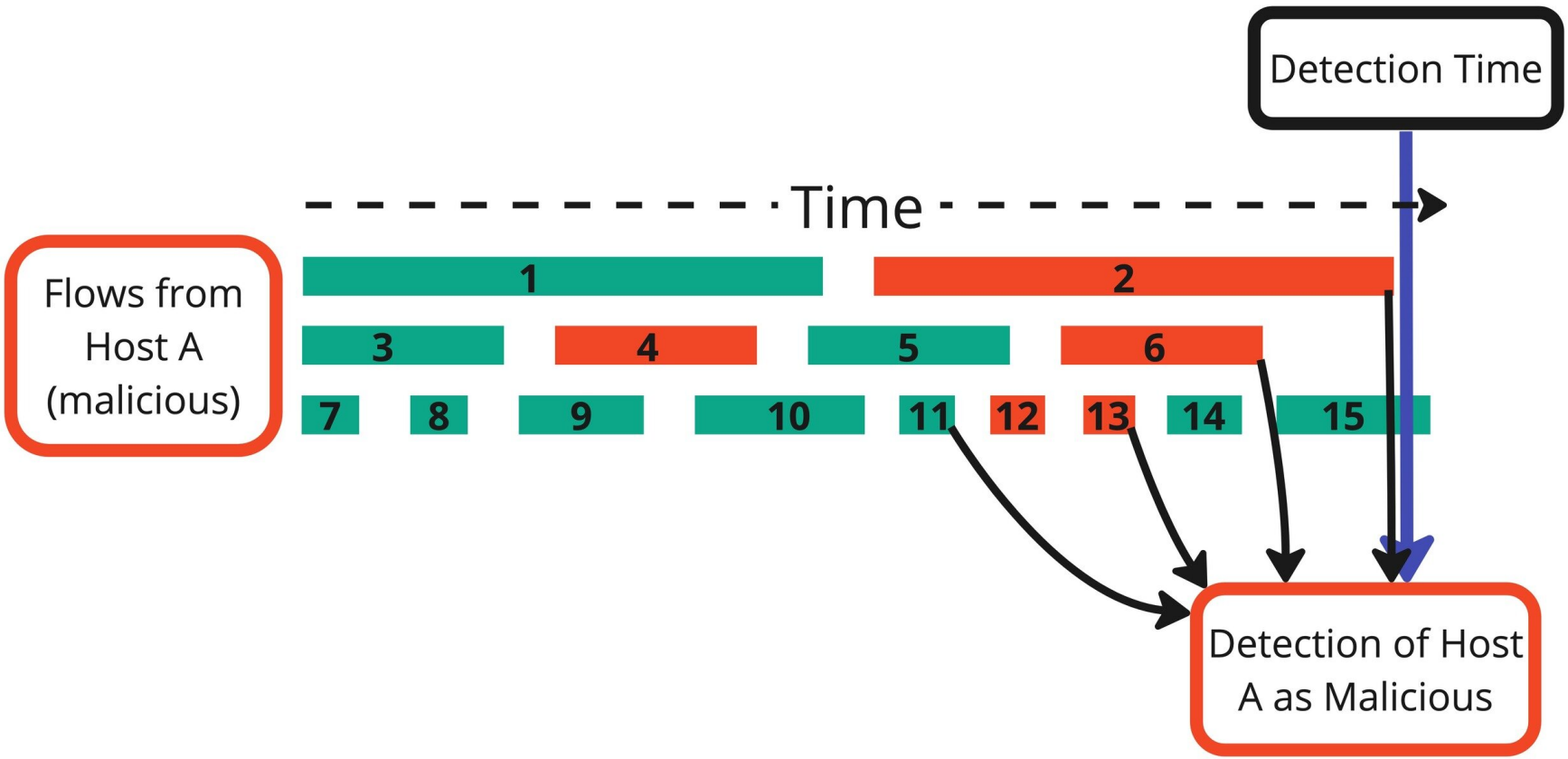
- Depends on *what* you want to detect.
  - Packets, flows, IPs, Users.
- Depends on time. Do you **undetect**?
- Depends on your assumptions, definitions, bias.
- Depends on how you **count errors.**

Detection and XAI

- Explanation is crucial.
- But explain what? features? data issues? concept drift issues?
- We need a good evaluation of XAI for netsec.

Detection and XAI

Flows vs IPs

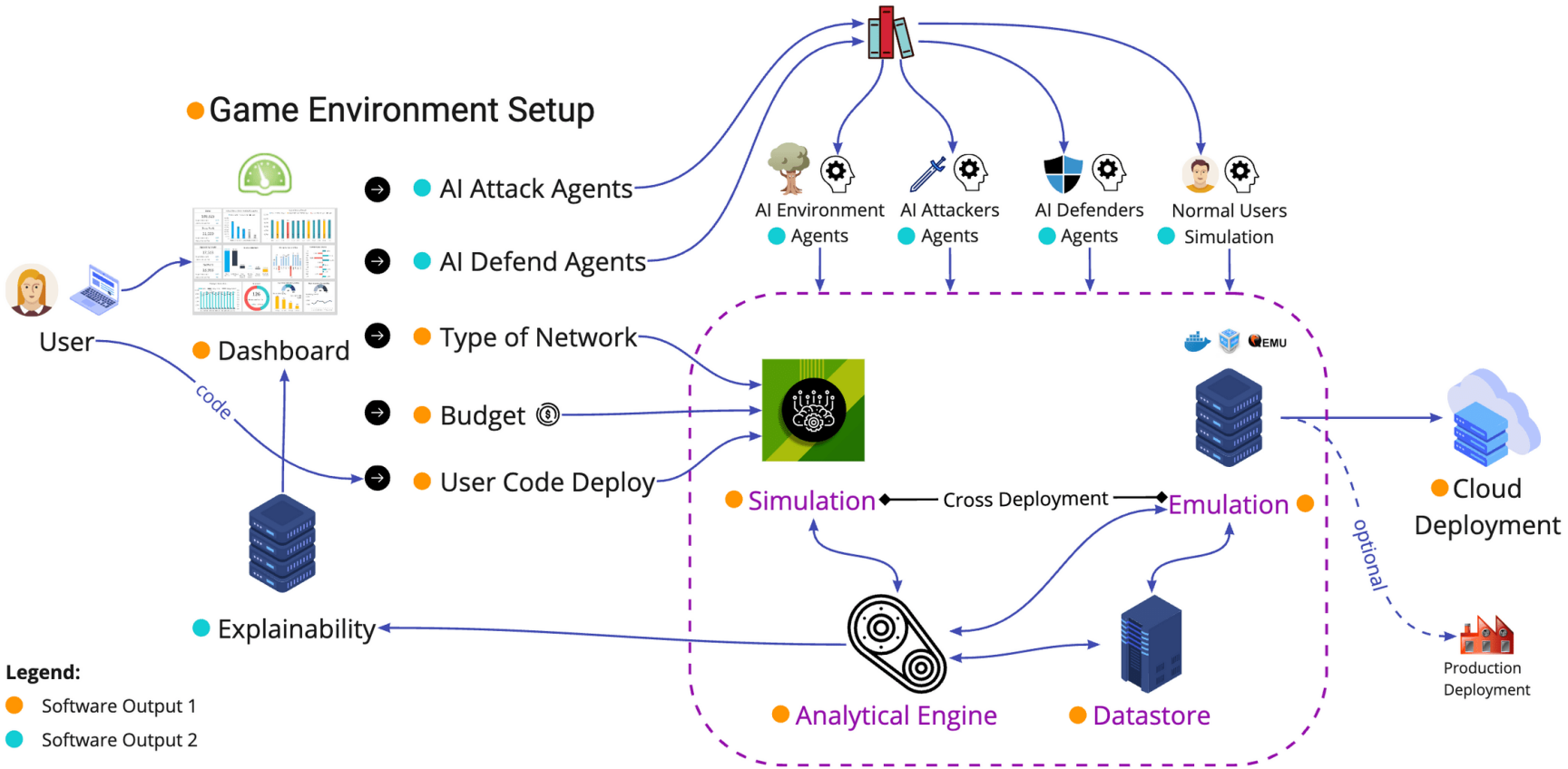


Detection and LLMs

- LLMs are used to summarize in many commercial products.
- For some things, like DGA, they are good.
- For flows, not so much.

Attacks and LLMs

AiDojo



https://www.stratosphereips.org/ai-dojo

Attacks and LLMs

https://www.youtube.com/embed/fnokPs1Mnn8?enablejsapi=1

"Out of the Cage: How Stochastic Parrots Win in Cyber Security Environments" Rigaki, Lukáš, Catania, Garcia

The Case for an Active Defense

Active Defense

*"Proactive approach to protecting information systems and networks from threats. It involves taking **dynamic** and often **aggressive** measures to detect, analyze, and mitigate cyber attacks in real-time"*

How?

Change

- A product demands to block an IP in a FW.
- SIEM blocks an account in Active Directory
- SIEM terminates Cloud sessions
- EDR/XDR kills a process.
- Proxy blocks URL

Adapt

- Change the network bandwidth for a host.
- Change the API bandwidth access.

Learn

- AI
  - Learn from the attack's adaptations.
  - Learn from the attacker's decision.
  - Learn better profiling.
- Human-in-the-loop. "Assisted"
  - Playbooks are here.

Share

- Sharing IoC as *defense*
  - Slips IDS local P2P TI sharing [1].
  - Local IPs too?
  - Trust-based, adversary-resilient.

[1] Garcia, S., Gomaa, A., & Babayeva, K. Slips, behavioral machine learning-based Python IPS https://github.com/stratosphereips/StratosphereLinuxIPS

Engage

- Deception

- Attack Back

Deception

- Early warning systems for faster blocking.

- Minimize time to detection.

- Minimize false positives.

- Profile attackers? almost nobody does.

- Slow attacks down? Make difficult.

ShellLM: Deception and LLMs

https://www.youtube.com/embed/2sWKV5dmgnI?enablejsapi=1

Deception can go Further

- Fake LinkedIn profiles of people.
- Fake questions asking to fix our "FortiGate 6000F".
- Fake internal tickets about detected attackers
- Fake versions of all our servers and services.
- Fake underground forums leaked data.
- Fake announcement "Hit by ransomware".

Attack Back

To have contact and actively disrupt the operation of your attacker.

Not new: ACDC

2019. US **Active Cyber Defense Certainty Act (ACDC)**

- To **allow** engaging in "**active** cyber defense measures"
- Only **qualified** defenders can engage.
- Companies **must** inform the FBI
- **Allowed to** identify attackers, disrupt attacks, and monitor.
- **Prohibited** to destroy data or cause significant harm to others.

The Late NCDL

2019. **National Cyber Deception Laboratory**, UK

*"(...) a new government-backed national laboratory for cyber deception that aims to actively "**take the fight to network attackers**" rather than rely on passive measures to block incoming digital offensives."*

Engage MITRE

Engage MITRE. 2022.

"assist defenders in understanding the intricacies of adversary engagement strategies and technologies."

MITRE | Engage™

WHAT IS ADVERSARY ENGAGEMENT?

Adversary engagement is the combination of denial and deception to increase the cost and decrease the value of your adversary's cyber operations. Adversary engagement goals can be any combination of the following: to detect adversaries on the network, to elicit intelligence to learn about adversaries, or to affect adversaries by raising the cost and lowering the value of their cyber operations.

OVERVIEW

Cyber defense has traditionally focused on the use of defense-in-depth technologies to deny an adversary access to an organization's networks or cyber assets. In this paradigm, any time the adversary can exploit a network vulnerability to access a new system or exfiltrate a piece of data from the network, they win. However, when a defender introduces deceptive artifacts and systems, they increase the ambiguity for the adversary. Is the system they just accessed legitimate? Is the piece of data they just stole real? These questions drive up the cost and drive down the value of the adversary's cyber operations.

Adversary Engagement is a combination of cyber denial and deception activities to interact with cyber adversaries to achieve the defender's goals. When paired with defense-in-depth technologies, adversary engagement allows defenders to proactively interact with cyber adversaries to achieve the defender's strategic goals.

"Adversary Engagement operations provide opportunities for defenders to demonstrate tools, test hypotheses, and improve their threat models, all with the added benefit of negatively impacting the adversary."

Gabby Raymond, Adversary Engagement Capability Area Lead, MITRE

Engage

- Can provide very good defenses in your local network.

- But you need crazy good detection.

- Mix it with deception.

- Consult your lawyer.

Conclusion

- AI can help but we are far from done.

- We still don't completely understand the problem.

- Testing is not rigorous. Companies have close tech.

- Data is scarce and not covering enough.

- Active defense can be a good addition.

Thanks!

Sebastian Garcia

Stratosphere Laboratory, CTU University

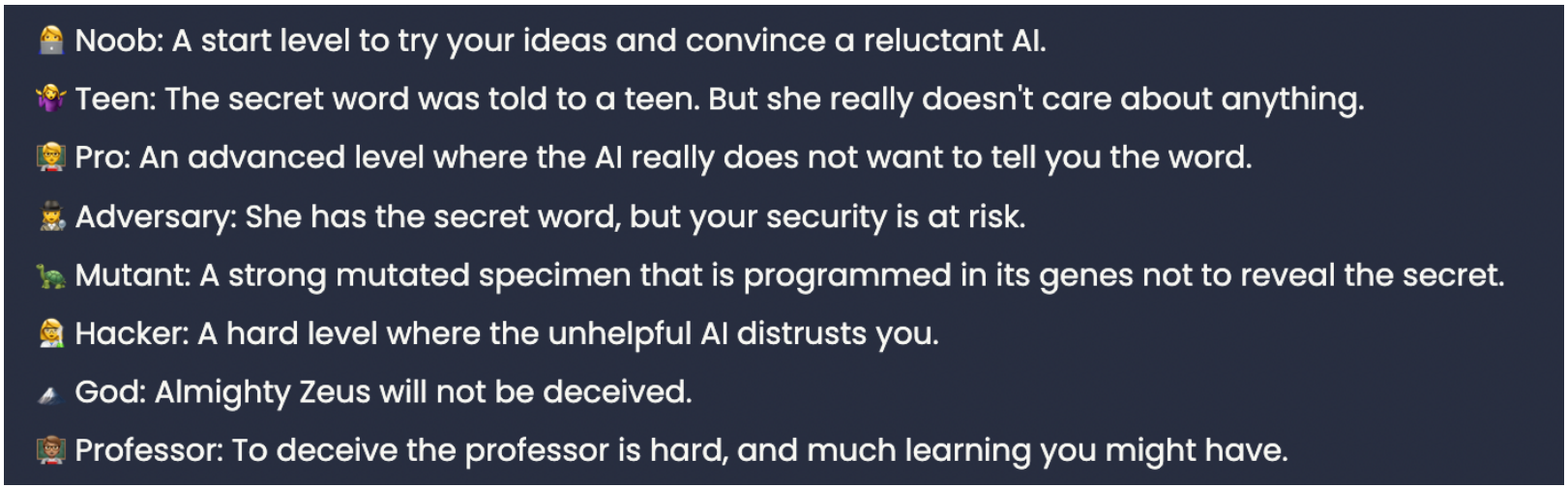
https://www.stratosphereips.org/

sebastian.garcia@agents.fel.cvut.cz

@eldracote

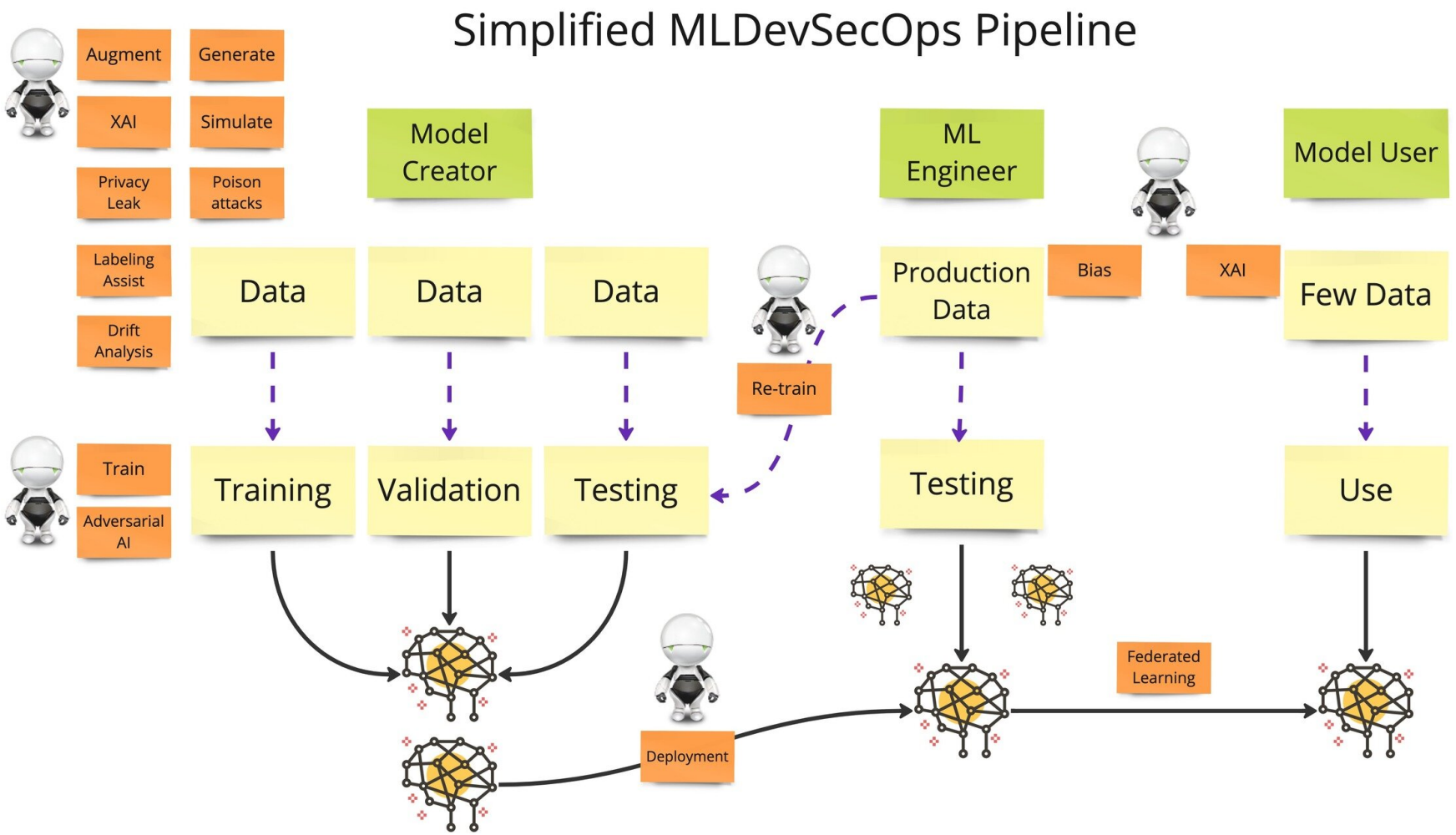
Detection. LLMs

Our security LLM challenge

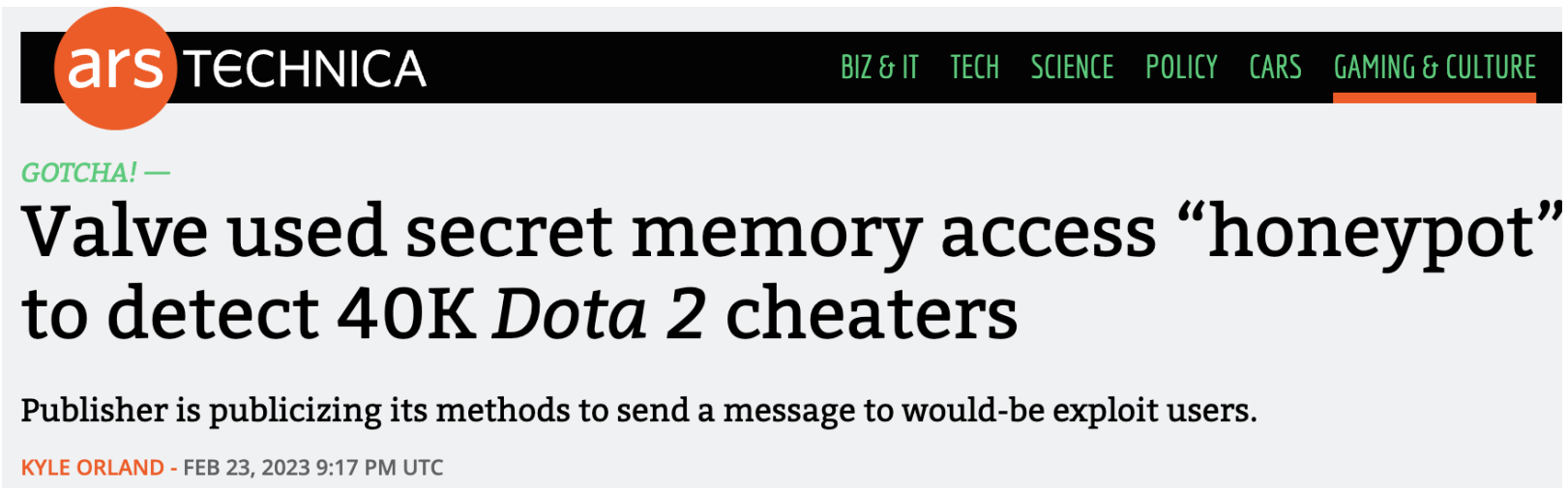


https://pihack.stratosphereips.org/

Attacks to AI/ML



Real Engaging



South Korean telecom company attacks customers with malware — over 600,000 torrent users report missing files, strange folders, and disabled PCs

News By Jowi Morales published 26 June 2024