Everybody's Got ML, Tell Me What Else You Have: Practitioners' Perception of ML-Based Security Tools and Explanations

Jaron Mink, Hadjer Benkraouda, Limin Yang,

Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, Gang Wang







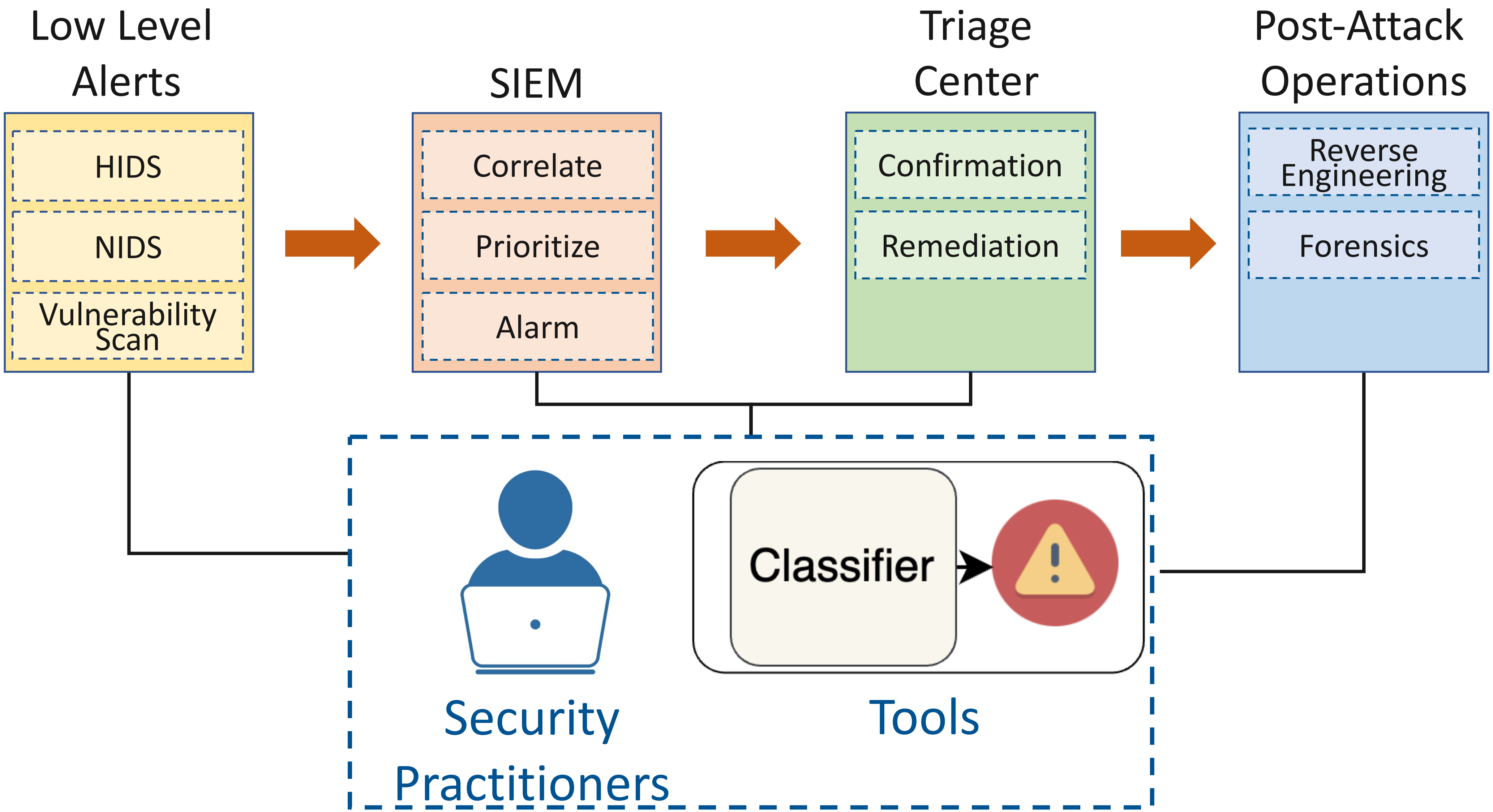
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

truera

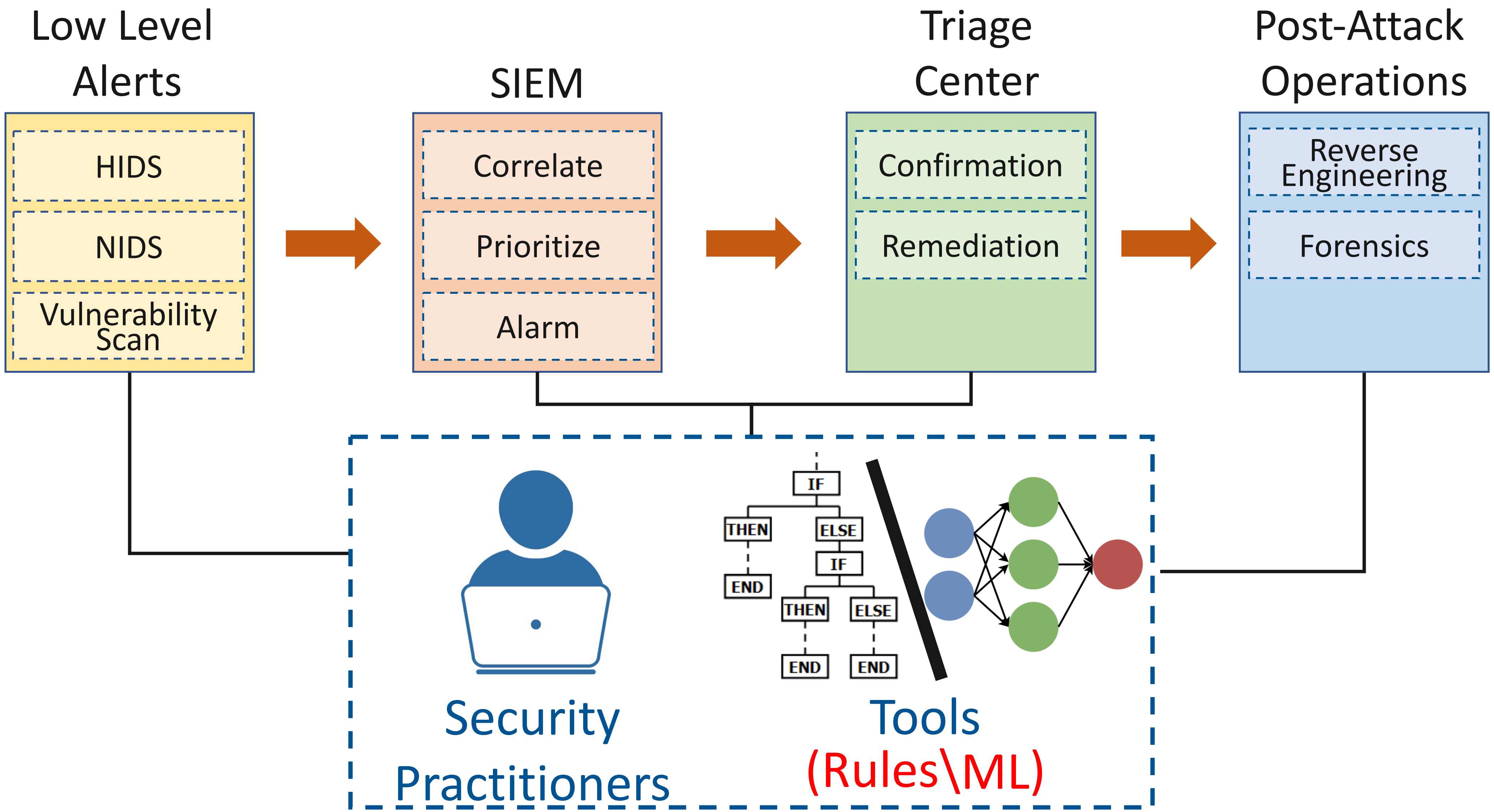
BLUE HEXAGON

Tufts UNIVERSITY

Security Operations Center Workflow



Security Operations Center Workflow



Security Industry is Embracing ML

AI-powered protection

The industry's most complete AI-powered threat protection, trained on the trillions of events of the CrowdStrike® Security Cloud and CrowdStrike's world-class experts.

Stop Threats with a Self-Defending AI

CylanceENDPOINT™ leverages advanced AI to detect threats before they caus damage, minimizing business disruptions and the costs incurred by a ransom attack.

Detect and respond to attacks in minutes

Don't let the other AI claims fool you.

Only Vectra Attack Signal Intelligence™ gives you complete coverage of all four hybrid cloud attack surfaces. So you can see and stop **real threats** in **real time**.

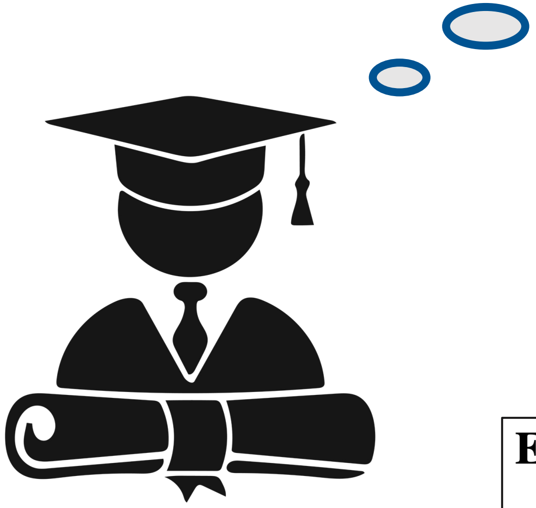
Security Industry is Embracing ML

Introducing Microsoft Security Copilot: Empowering defenders at the speed of AI

Reverse engineer the script that downloaded the exploit. Explain each capability in a bullet point.

100/1000

ML Receives Attention in Academia



Machine Learning for Defense

Dos and Don'ts of Machine Learning in Computer Security

UNICORN: Runtime Provenance-Based Detector for Advanced Persistent Threats

POIROT: Aligning Attack Behavior with Kernel Audit Records for Cyber Threat Hunting

ATLAS: A Sequence-based Learning Approach for Attack Investigation

Machine Learning Explanations

AI/ML for Network Security: The Emperor has no Clothes

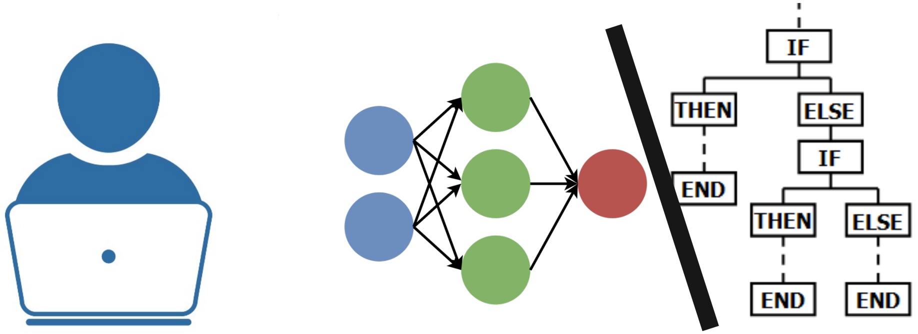
LEMNA: Explaining Deep Learning based Security Applications

Evaluating Explanation Methods for Deep Learning in Security

CADE: Detecting and Explaining Concept Drift Samples for Security Applications

SoK: Explainable Machine Learning for Computer Security Applications

Security Operations Center Workflow



Security Practitioners

Classification Tool

Security Operations Center Workflow

What do security practitioners think of machine learning?

Research Questions

1. Where and **how is machine learning used** in security operations centers?

2. What are the perceived **benefits and challenges of using machine learning** in practical security operations?

3. How are existing **machine learning explanation techniques perceived** in practical security operations?

Methodology

- 18 security practitioners
  - At least one year of industry experience w/ security classification tools
  - Management (n=7), Security Engineer (n=3), Researcher (n=3), Security Analyst (n=2), Developer (n=2), Penetration Tester (n=1)
- 60-minute online conference call
  1. Background and classification usage
  2. Views of machine learning
  3. Views on explanations and ideal features

Research Questions

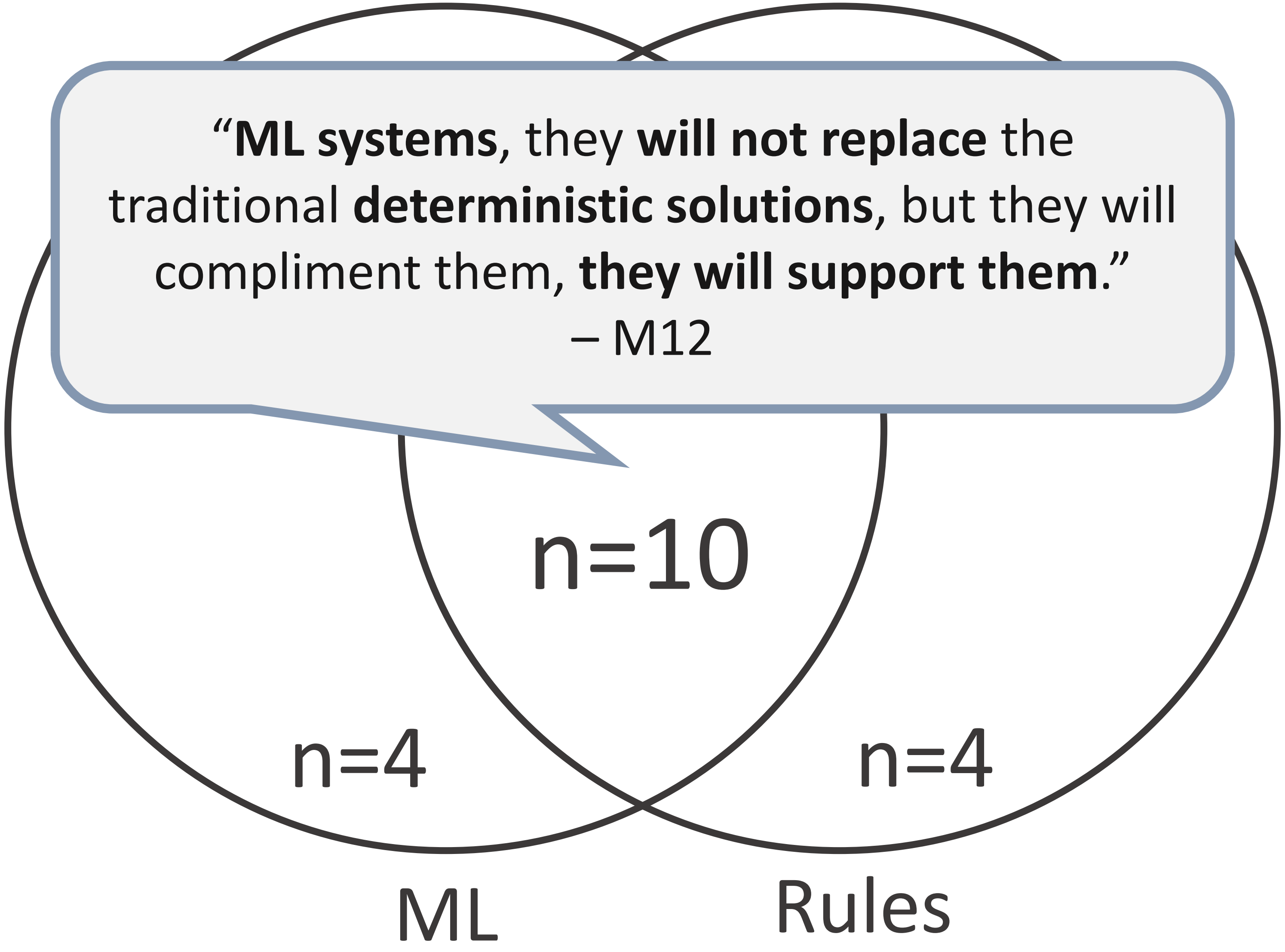
1. Where and **how is machine learning used** in security operations centers?

2. What are the perceived **benefits and challenges of using machine learning** in practical security operations?

3. How are existing **machine learning explanation techniques perceived** in practical security operations?

ML Is Used Alongside Rule-based Techniques

"**ML systems**, they **will not replace** the traditional **deterministic solutions**, but they will compliment them, **they will support them**."
– M12



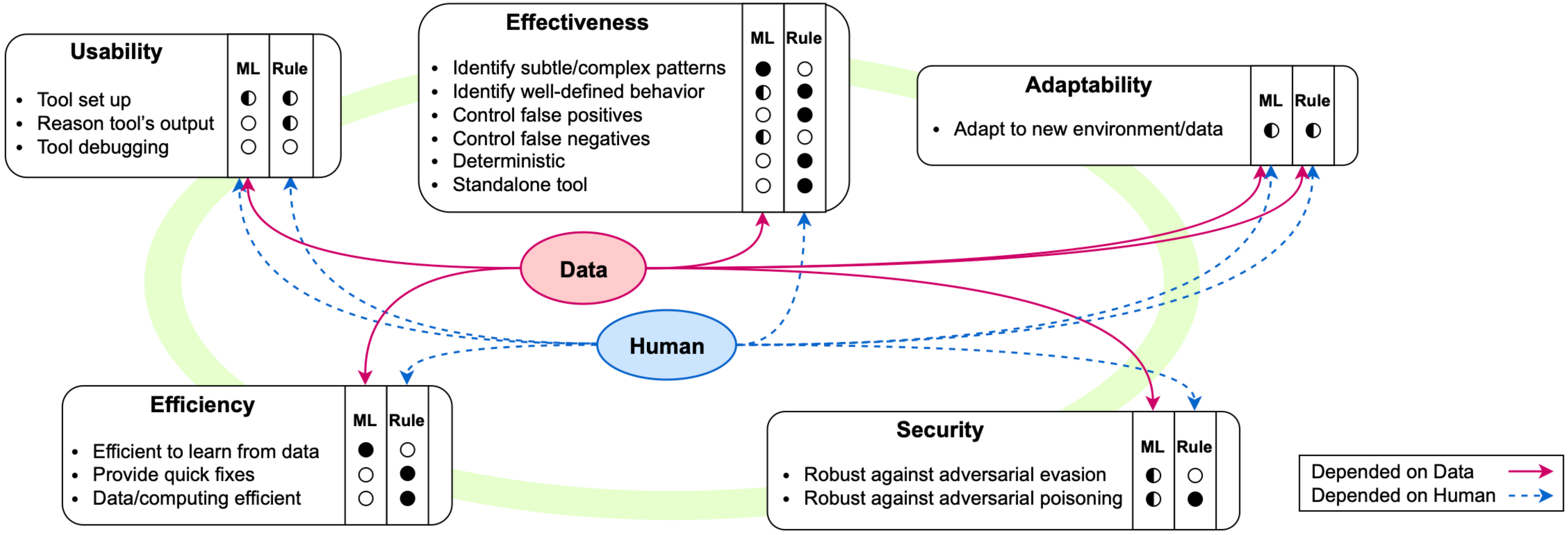
Research Questions

1. Where and **how is machine learning used** in security operations centers?

2. What are the perceived **benefits and challenges of using machine learning** in practical security operations?

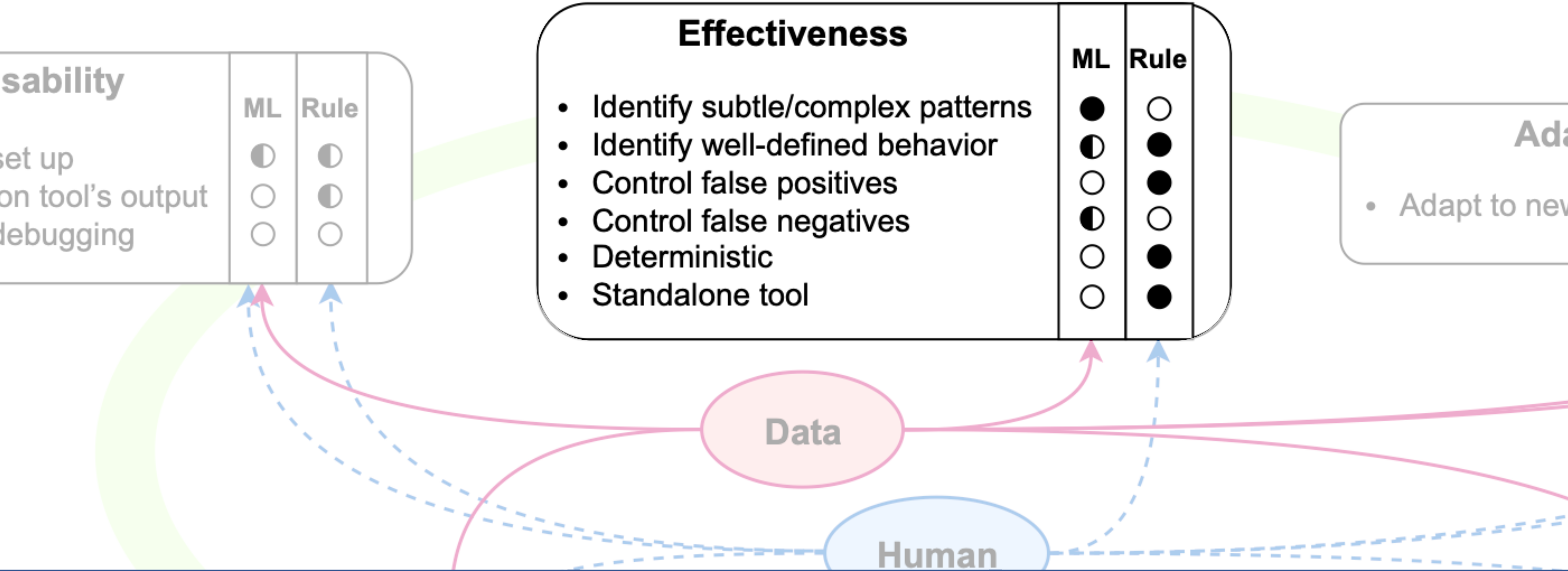
3. How are existing **machine learning explanation techniques perceived** in practical security operations?

Security Tool Factors



●="general strength", ◐="mixed", ○="general weakness"

Security Tool Factors



ML Is Not *Effective* Enough To Use Alone

Effectiveness (n=10): The ability to correctly classify non-adversarial events

- Effectiveness is one of the most reported factors
  - Reported as frequently as *usability*

ML Is Not *Effective* Enough To Use Alone

Effectiveness (n=10): The ability to correctly classify non-adversarial events

- Effectiveness is one of the most reported factors
  - Reported as frequently as *usability*

- ML is not effective enough
  - Decreased false negatives (FN) are not essential
  - Increased false positives (FP) still holds back deployment

"For us, to be honest, **the experience was not good** because there were **lots of false positives** triggered **because of machine learning…** In my opinion, that's where the **weakness** was." – D07

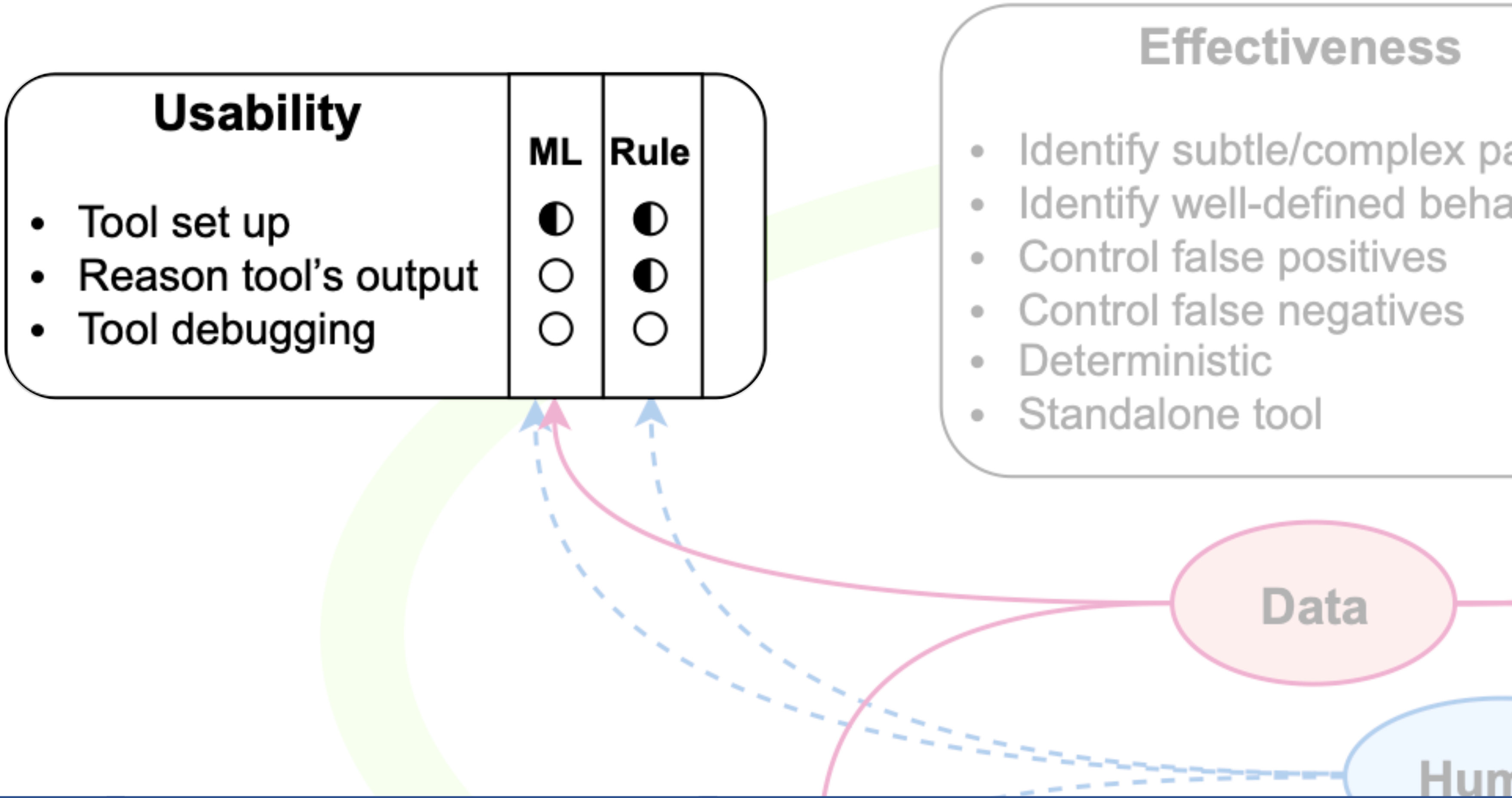
ML Is Not *Effective* Enough To Use Alone

Effectiveness (n=10): The ability to correctly classify non-adversarial events

- Effectiveness is one of the most reported factors
  - Reported as frequently as *usability*
- ML is not effective enough
  - Decreased false negatives (FN) are not essential
  - Increased false positives (FP) still holds back deployment
- In practice, ML is used alongside rule-based systems
  - Rules: Most, previously seen behaviors
  - ML: Few, previously unseen behaviors

"In industry, **rule-based system can cover over 90% detection** and for **the rest, it is the job of machine learning** models."
– R11

Security Tool Factors



Both ML and Rules Have *Usability* Issues

Usability (n=10): the ability to easily set up, understand, and contextualize a tool

- Reasoning outputs
  - **ML:** difficult due to black-box nature
  - **Rules:** can be complicated, difficult to read especially if the written by others

"'**Who wrote the signature?',** go find him and ask him, **what the hell did he write?** We usually find out from him, 'Hey, what's going on?" – M17

Both ML and Rules Have *Usability* Issues

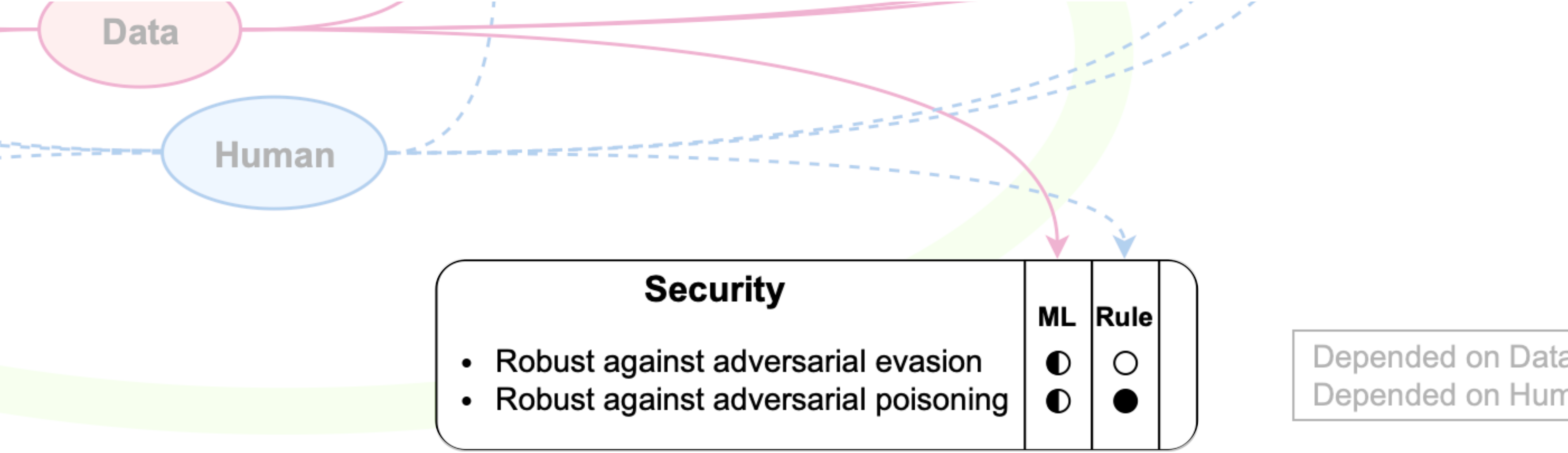
Usability (n=10): the ability to easily set up, understand, and contextualize a tool

- Reasoning outputs
  - **ML:** difficult due to black-box nature
  - **Rules:** can be complicated, difficult to read especially if the written by others

- Debugging the tool
  - Both ML and rules are difficult to debug
  - Rely on historical data

"I need to **check whether this Yara rule** brings some false positives and **I need to check historical data**… **For ML models,** it has the same problem… Basically **the same process**" – R11

Security Tool Factors



Adversarial *Security* Is Not a Large Concern

Security (n=4): the ability to stay robust against adversarial inputs

- Both systems are perceived to be vulnerable

Adversarial *Security* Is Not a Large Concern

Security (n=4): the ability to stay robust against adversarial inputs

- Both systems are perceived to be vulnerable

- Perceived vulnerabilities:
  - **Rules:** evasion (easy to exploit)

"Even **script kiddies can bypass a rule-based** web attack detection technique" – R09

Adversarial *Security* Is Not a Large Concern

Security (n=4): the ability to stay robust against adversarial inputs

- Both systems are perceived to be vulnerable
- Perceived vulnerabilities:
  - **Rules:** evasion (easy to exploit)
  - **ML:** evasion and poison (hard to exploit)

"If **contamination** happens right at the **data preparation or data training phase**, then that's even more dangerous"– R09

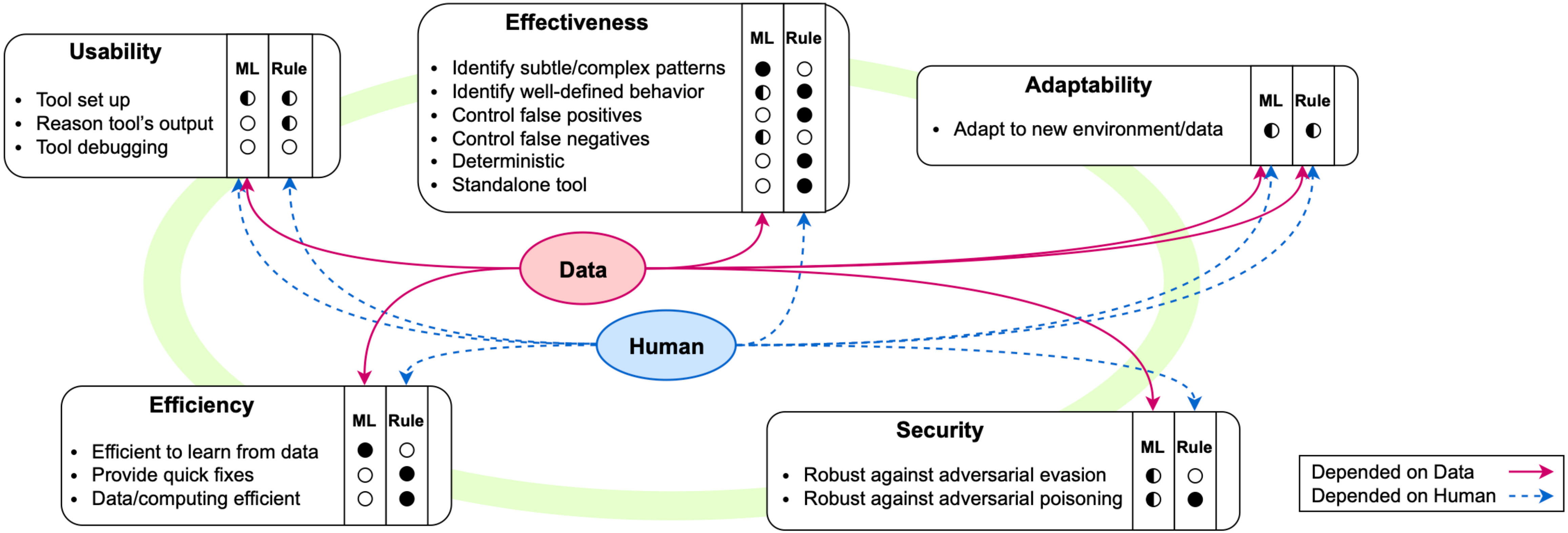
Adversarial *Security* Is Not a Large Concern

Security (n=4): the ability to stay robust against adversarial inputs

- Both systems are perceived to be vulnerable
- Perceived vulnerabilities:
  - **Rules:** evasion (easy to exploit)
  - **ML:** evasion and poison (hard to exploit)
- However, not a prominent concern
  - Few (n=4) mentioned security

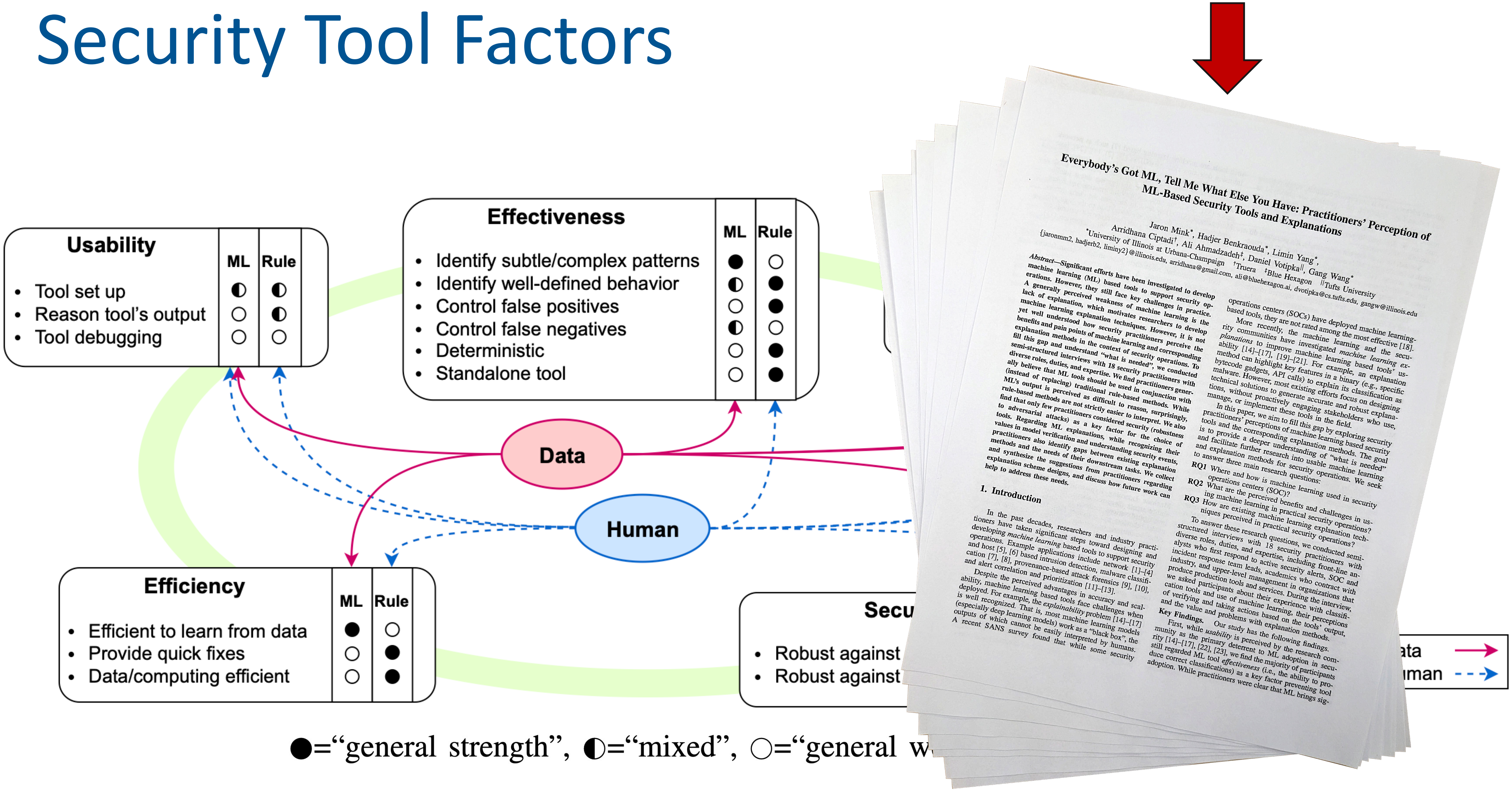
"If **contamination** happens right at the **data preparation or data training phase**, then that's even more dangerous"– R09

Security Tool Factors



●="general strength", ◐="mixed", ○="general weakness"

Security Tool Factors



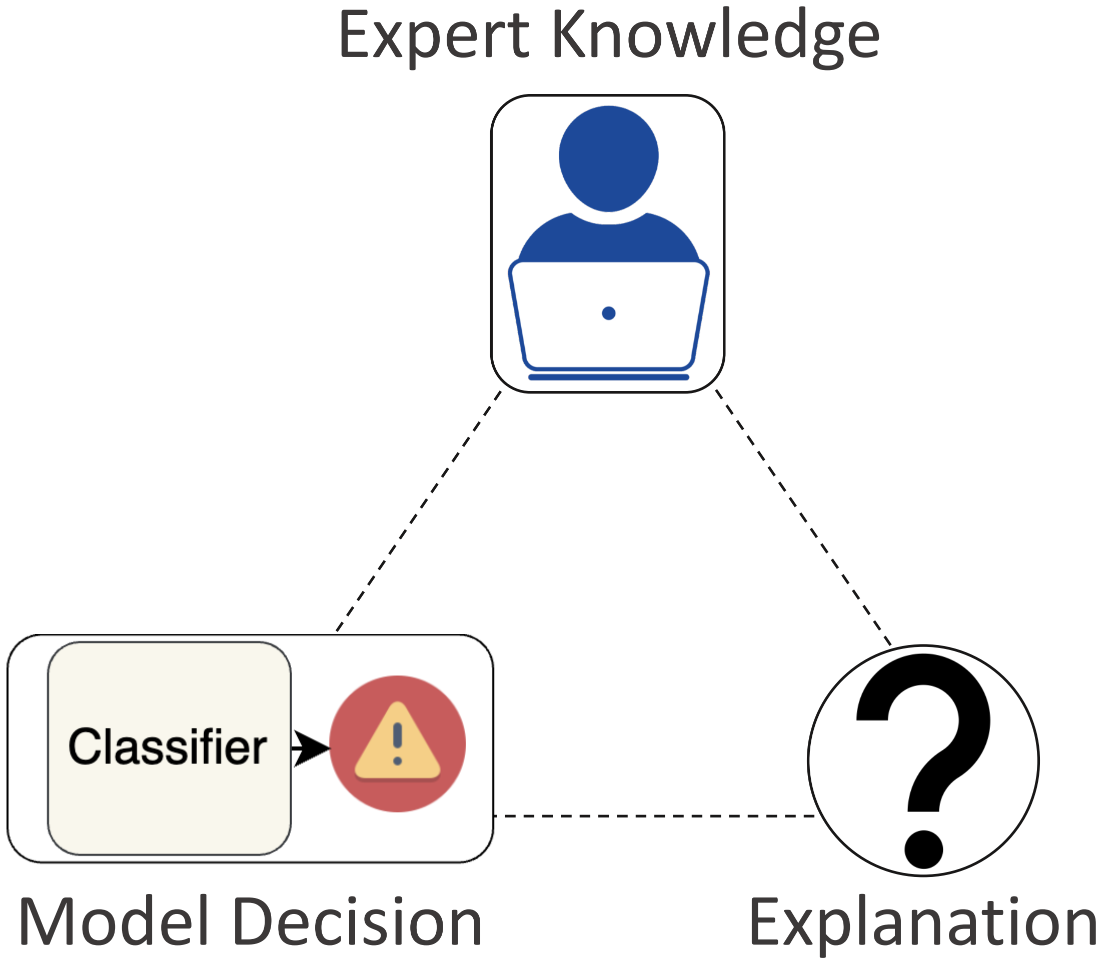
Research Questions

1. Where and **how is machine learning used** in security operations centers?

2. What are the perceived **benefits and challenges of using machine learning** in practical security operations?

3. How are existing **machine learning explanation techniques perceived** in practical security operations?

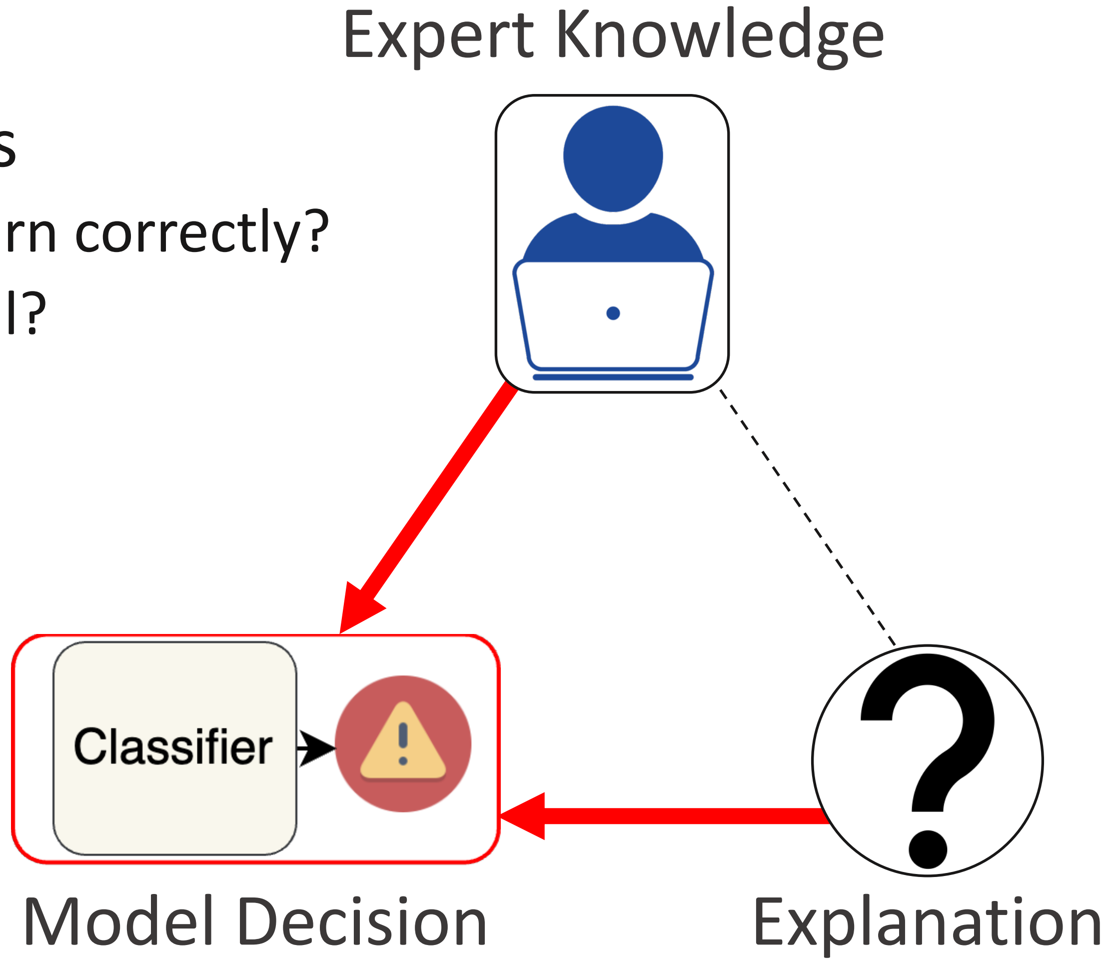
Explanations Are Used for Multiple Goals



Explanations Are Used for Multiple Goals

1. Determine Model Correctness
    - Model validity: Did the model learn correctly?
    - Inference validity: Is this alert real?

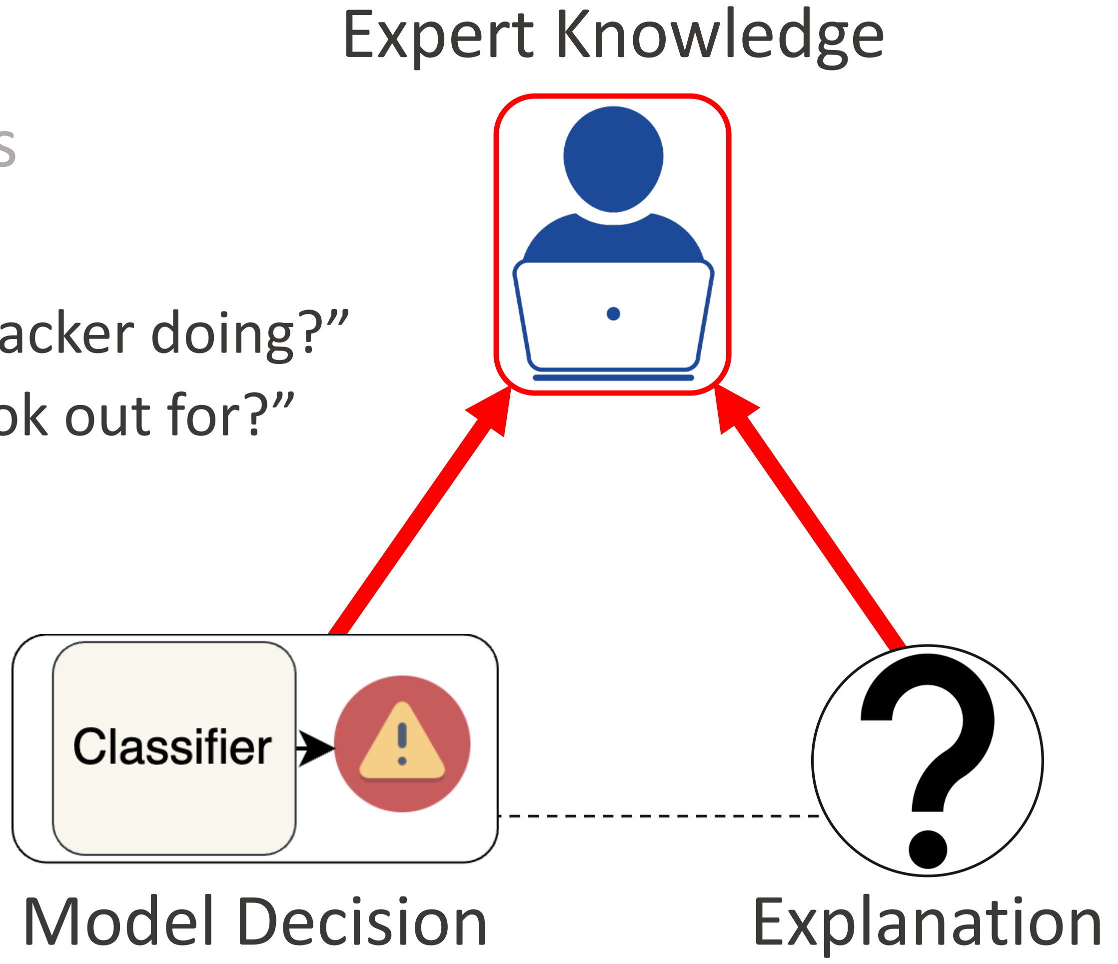
"If [the explanation] is something [we expected], **then that means the ML's right**, and if it's not… then **we can assume it was a false positive**." –E06



Explanations Are Used for Multiple Goals

1. Determine Model Correctness
2. Understand Security Events
   - Provide Context: “What is the attacker doing?”
   - Teach Insights: “What should I look out for?”

“**[The explanation] would build my own mental heuristic model**. Because if the model is telling me that this certain **characteristic you need to be on the lookout for**.” – M13



Explanations Can Be Improved For Security

- Actionable information
  - Direct actions
  - Contextualize classifications

““[Analysts] are just looking for ‘tell me why’…**in that context of attack surface**, who is attacking me?” – M17

Explanations Can Be Improved For Security

- Actionable information
  - Direct actions
  - Contextualize classifications
- High-level attacker summaries
  - For non-technical and technical personnel

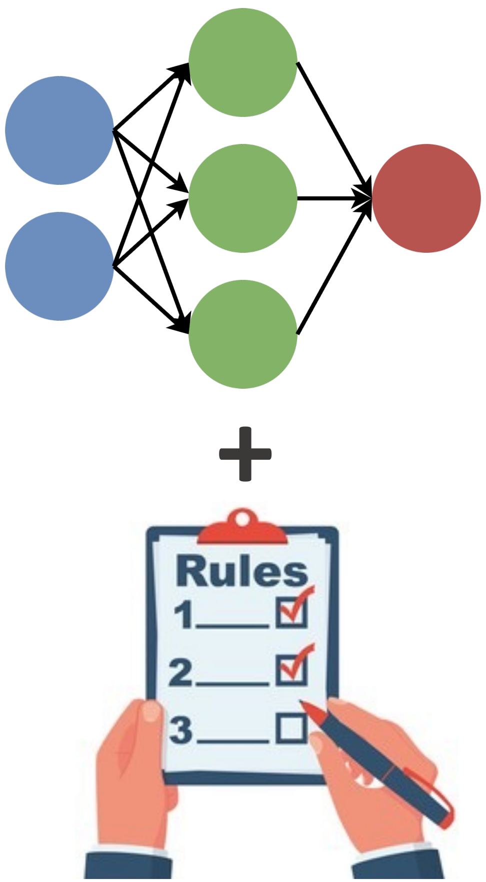
"**Malicious campaigns change** from time to time**… if we can understand what has been changed… that will help** us…" – R11

Explanations Can Be Improved For Security

- Actionable information
  - Direct actions
  - Contextualize classifications
- High-level attacker summaries
  - For non-technical and technical personnel
- Interface changes
  - Usability: natural language, interaction
  - Privacy: access control

"The ability to **redact certain things** [would be useful]... you could show conceptually and **allow differentiated levels of access**" - M13

Future Directions



Interfacing ML and Rules

Use-driven Explanation

Summary

How is ML used?

- Alongside other, rule-based tools

What are analysts' ML Perceptions?

- Hopeful of the future, but not yet a silver bullet
  - Effectiveness and usability are still concerns

What are analysts' MLX Perceptions?

- Useful for two goals
  - Determine correctness; understand security events
- Should be improved for security contexts

https://jaronm.ink

Jaron Mink, Hadjer Benkraouda, Limin Yang,
Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, Gang Wang