On Managing Vulnerabilities in AI/ML Systems

Hot Topics in the Science of Security, April 15, 2021

Jonathan M Spring, April Galyardt,
Allen D Householder, Nathan VanHoudnos

Published at NSPW 2020,
https://dl.acm.org/doi/10.1145/3442167.3442177

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

This material is based upon work funded and supported by the Department of Homeland Security under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense.

DM21-0306

What if **flaws** in machine learning (**ML**) were **assigned** Common Vulnerabilities and Exposures (CVE) identifiers (**CVE-ID**s)?

Flaw

“A vulnerability is a set of conditions or behaviors that allows the violation of an explicit or implicit security policy. Vulnerabilities can be caused by software defects, configuration or design decisions, unexpected interactions between systems, or environmental changes.”

https://vuls.cert.org/confluence/display/CVD/1.2.+CVD+Context+and+Terminology+Notes

“Flaw”

“A vulnerability is a set of conditions or behaviors that allows the violation of an explicit or implicit security policy. Vulnerabilities can be caused by software defects, configuration or design decisions, unexpected interactions between systems, or environmental changes.”

https://vuls.cert.org/confluence/display/CVD/1.2.+CVD+Context+and+Terminology+Notes

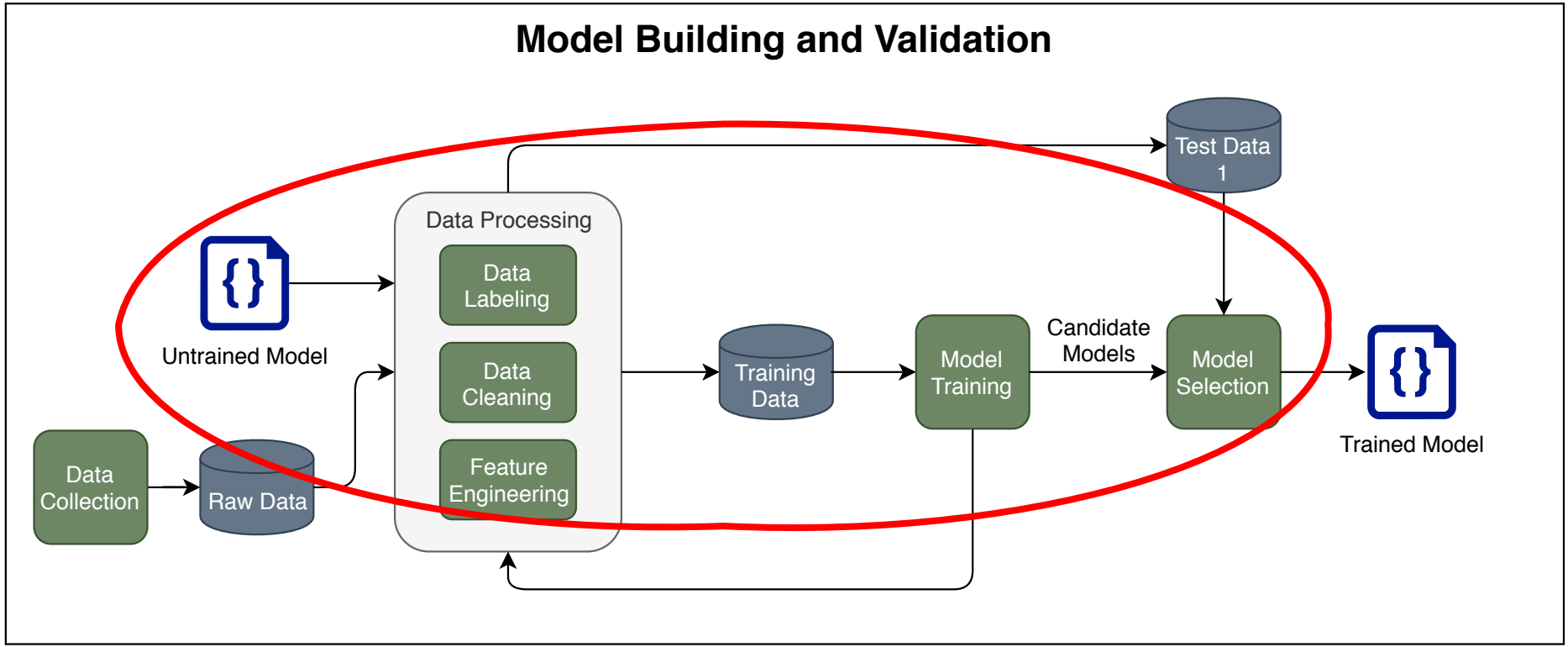
“Flaw”

“A vulnerability is a set of conditions or behaviors that allows the violation of an explicit or implicit security policy.
Vulnerabilities can be caused by software defects, configuration or design decisions, unexpected interactions between systems, or environmental changes.”

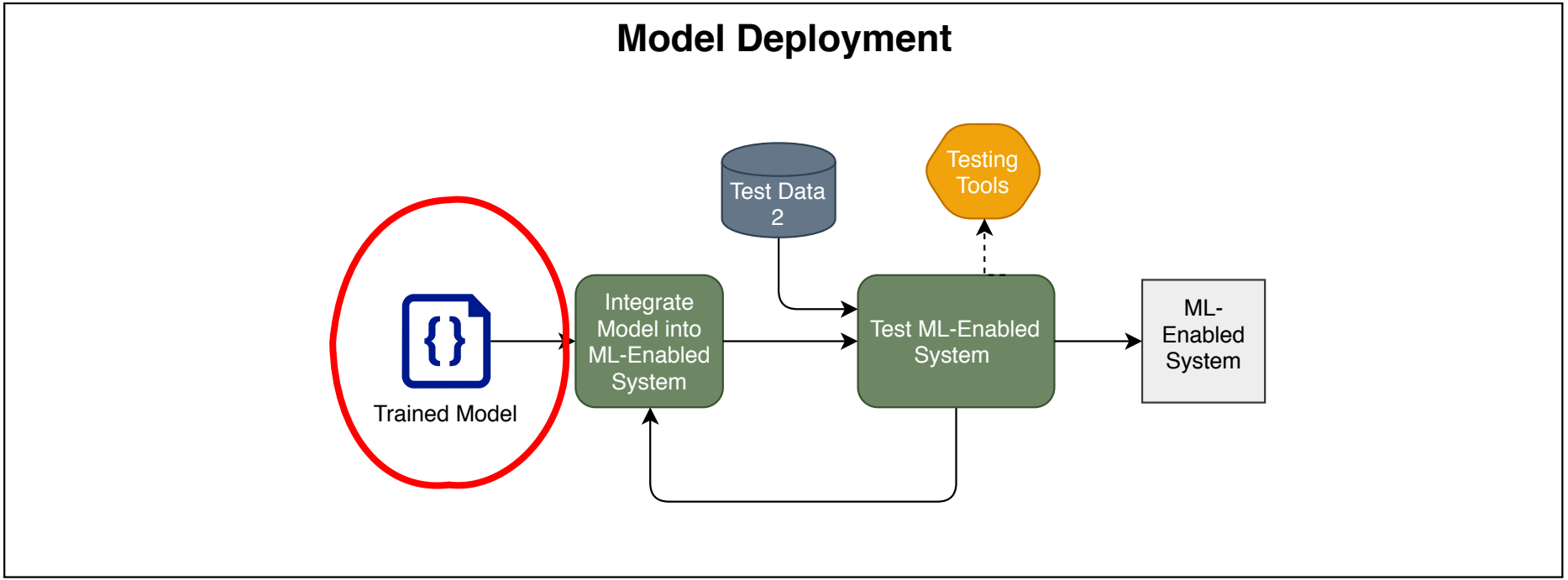
https://vuls.cert.org/confluence/display/CVD/1.2.+CVD+Context+and+Terminology+Notes

Assigned to ML algorithm or model object?

ML algorithm or ML model object



ML algorithm or ML model object



CVE-IDs are:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

CVE-IDs are:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

CVE-IDs may:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

CVE-IDs are:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

CVE-IDs are:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

CVE-IDs are:

- Assigned to implementation vulnerabilities in products
- Assigned to protocol vulnerabilities
- Tag any instance of a vulnerable product

CVE-IDs are not:

- Assigned to individual instances of vulnerabilities
- Assigned to categories of vulnerabilities (CWE)
- Assigned to configuration errors

What would change in vulnerability management if CVE-IDs were assigned to ML algorithms? ML model objects?

Vulnerability Management† and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

† https://www.first.org/standards/frameworks/csirts/csirt_services_framework_v2.1#7-Service-Area-Vulnerability-Management

Vulnerability Management and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Algorithms

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management† and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

† https://www.first.org/standards/frameworks/csirts/csirt_services_framework_v2.1#7-Service-Area-Vulnerability-Management

Vulnerability Management and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Vulnerability Management and ML Model Objects

1. Vulnerability discovery / research
2. Vulnerability report intake
3. Vulnerability analysis
4. Vulnerability coordination
5. Vulnerability disclosure
6. Vulnerability response

Thanks for your time.

Questions?

Contact: spring AT cmu<>edu
or https://kb.cert.org/vuls/report/