

VLAI: A RoBERTa-Based Model for Automated Vulnerability Severity Classification

Vulnerability Forecasting Technical Colloquium - Cambridge, United Kingdom

<https://arxiv.org/abs/2507.03607>

Cédric Bonhomme - cedric.bonhomme@circl.lu

September 26, 2025

CIRCL <https://www.circl.lu>

1. Origin of Vulnerability-Lookup
2. Where the Data Comes From
3. From Data to Datasets
4. From Datasets to Models

Origin of Vulnerability-Lookup

Who is behind Vulnerability-Lookup?



Vulnerability-Lookup¹ is an Open Source project led by **CIRCL**.
It is co-funded by **CIRCL** and the **European Union**².
Used by many organisations including CSIRTs and ENISA (EUVD).



vulnerability
-lookup

¹<https://www.vulnerability-lookup.org>

²<https://github.com/ngsoti>

What is Vulnerability-Lookup?

- A unified place to **search**, **triage**, and **track** software and product vulnerabilities from different sources.
- Brings together vulnerability information, correlation of identifiers (e.g., CVE, GHSA, OSV), references, timelines, and risk signals in one view.
- Designed for **CSIRTs**, **SOCs**, **vulnerability managers**, and **developers** in mind.
- **Web UI first** with strong **API** and automation options.
- Support a complete **CVD process management**³ along with the ability to fork vulnerability information⁴. Distributed GNA directory.

³CNA Program, GCVE GNA Publication

⁴Implemented before the GNA initiative!

Origin of Vulnerability-Lookup

- `cve-search`⁵ is an open-source tool initially developed in late 2012, focusing on maintaining a **local** CVE database.
- `cve-search` is widely used as an **internal** tool.
- The design and scalability of `cve-search` are limited. Our operational public instance has reached a hard limit of 20,000 queries per second.
- Vulnerability sources have **diversified**, and the **NVD CVE is no longer the sole source** of vulnerability information.

⁵<https://github.com/cve-search/cve-search>

Initial Challenges

- **Volume of data:** Handling a substantial dataset and heavy network traffic, currently over 1,723,037 security advisories and more than 130,000 sightings collected in one year ⁶.
- **Flexibility:** Balancing ongoing development with legacy issues while designing a future-proof architecture. It's complex and yes, sometimes chaotic⁷.
- **Robustness:** Validating data even when external entities don't comply with their own JSON schemas. It's not always pretty.
- **Fast lookup:** Rapidly correlating identifiers across **diverse sources**, including unpublished advisories.

⁶The first sighting on Exploit-DB dates back 26 years.

⁷We enjoy challenges, especially when they lead to practical solutions.

Ongoing Challenges and Development

- **CPE fragmentation:**⁸ Tackling the fragmentation of CPEs (e.g., `cpe:/a:oracle:java` vs. `cpe:/a:sun:java`) by introducing *Organizations* as unified containers.
- **CVD process:** Building an open-source tool to support the Coordinated Vulnerability Disclosure process.⁹
- **Vulnerability numbering:** Enabling a new distributed approach through the Global CVE Allocation System.¹⁰
- **Scoring vulnerabilities:** Aggregating a **large volume** of observations from diverse advisory types to improve vulnerability scoring. Automatic **weighting** of sightings by source or type (e.g., exploited). Automatic **detection of the type** not depending on the source.

⁸Well, another mess to clean up!

⁹Aligned with NIS 2 and the Cyber Resilience Act.

¹⁰<https://gcve.eu>

Where the Data Comes From

1. Origin of
Vulnerability-Lookup

2. Where the Data Comes From

1. Current Sources
2. Feeders

3. From Data to Datasets

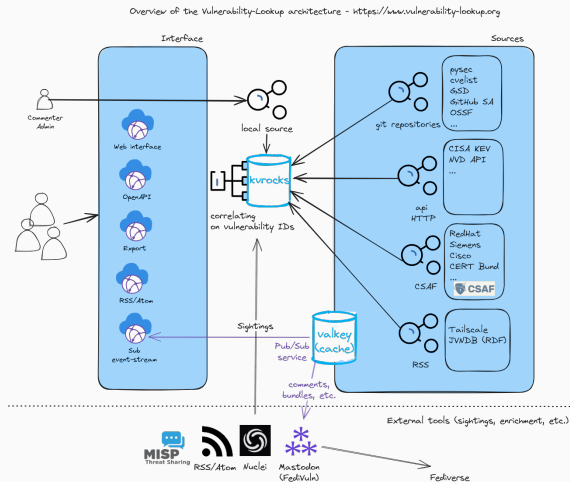
4. From Datasets to Models

It's a lot of sources!

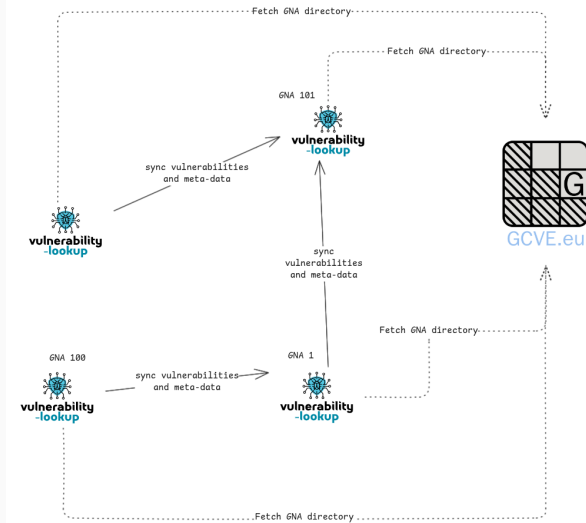
- **CVE Program** Git
- **NIST NVD CVE** API 2.0
- **Fraunhofer FKIE CVE** Git
- **GitHub Advisory Database** Git
- **Python Advisory Database** Git
- **Cloud Security Alliance - GSD** Git
- **VARIoT** API
- **GCVE** API
- **CSAF 2.0 (HTTP CSAF)**
CERT-Bund, Cisco, Siemens, Red Hat, Suse, OpenSuse, Microsoft, NCSC-NL, CISA, etc.
- **Japan - JVN DB** HTTP
- **China - CNVD** HTTP
- **CERT-FR** HTTP
- **Tailscale** RSS
- **CISA KEV** HTTP
- **EU KEV** HTTP
- **Growing...**

Vulnerability-Lookup High-Level Architecture

- Collects and aggregates vulnerability data.
- Preserves data integrity (never alters the original content).
- Harmonizes data using kvrocks.
- Correlates information across multiple sources (CVE, PySec, etc.).
- Provides both APIs and a Web interface.
- Supports advanced queries and filtering.



A Distributed Network



From Data to Datasets

1. Origin of Vulnerability-Lookup
2. Where the Data Comes From
3. From Data to Datasets
 1. Open Data Approach
 2. AI Datasets
4. From Datasets to Models

- **Open Data Initiative:** CIRCL's commitment to making data openly available.
- Consistent open approach applied across all our projects.
- Regularly updated JSON dumps ¹¹ and “AI” datasets ¹².
- Public, unauthenticated API access for Vulnerability-Lookup.

¹¹<https://vulnerability.circl.lu/dumps/>

¹²<https://huggingface.co/CIRCL/datasets>

The Messy Reality of Large Datasets

- Our experience with large datasets is not recent (Passive DNS¹³, BGP ranking¹⁴, MISP¹⁵, AIL¹⁶, Lookyloo¹⁷, etc.). And we learned from our past mistakes.
- Adapt to real-world conditions — avoid creating yet another format or standard.
- Deal with missing data, malformed JSON, and conflicting information.
- Tolerate unreliable or unstable remote servers (e.g., some CSAF providers).

¹³<https://www.circl.lu/services/passive-dns/>

¹⁴<https://github.com/D4-project/BGP-Ranking>

¹⁵<https://github.com/MISP>

¹⁶<https://github.com/ail-project>

¹⁷<https://github.com/Lookyloo>

- Turn messy data into structured, actionable insights.
- Link related vulnerabilities via enrichment, correlation, and crawling.
- Support the process with **VulnTrain**¹⁸.

¹⁸<https://github.com/vulnerability-lookup/VulnTrain>

Current datasets

Dataset	Size (rows)	Generation Time	Features
vulnerability-scores ¹⁹	641,547	10m45s	Descriptions (en), CVSS, CPE
vulnerability-CNVD ²⁰	122,546	1m31s	Descriptions (cn), CVSS
vulnerability-cwe-patch ²¹	883	210m	Descriptions (en), CWE, patches (commit id + url + full diff)

¹⁹<https://huggingface.co/datasets/CIRCL/vulnerability-scores>

²⁰<https://huggingface.co/datasets/CIRCL/Vulnerability-CNVD>

²¹<https://huggingface.co/datasets/CIRCL/vulnerability-cwe-patch>

From Datasets to Models

1. Origin of Vulnerability-Lookup
2. Where the Data Comes From
3. From Data to Datasets
4. From Datasets to Models
 1. Model generation workflow
 2. Examples
 3. Integration for inference

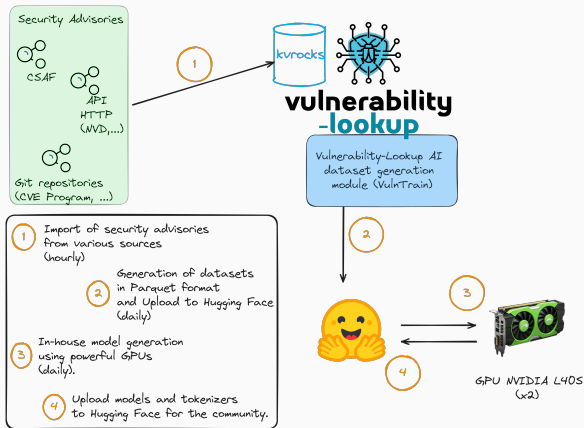
Why We Are Building AI Models

- CIRCL AI approach²²: we enhance existing solutions rather than replacing functional systems with NLP/ML/LLM solutions.
- AI-powered **enrichment** of vulnerability descriptions.
- Providing actionable insights to security experts when data is **missing or inaccurate** (e.g., severity, CWE, CPE information).
- We actively participate in collaborative research and development efforts, such as the EU-funded AIPITCH (AI-Powered Innovative Toolkit for Cybersecurity Hubs) project²³

²²<https://circl.lu/pub/ai-strategy/>

²³<https://www.science.nask.pl/en/research-areas/projects/12456>

Model generation workflow



- local training
- models are publicly shared
- regular update

Model	Size	Epochs	Accuracy	Training Time
Severity classification ²⁴	125M params	5	0.8289	6.72h
Severity classification (CNVD) ²⁵	102M params	5	0.7817	65.989m
CWE guessing ²⁶	125M params	36-40	0.875	30m

²⁴<https://huggingface.co/CIRCL/vulnerability-severity-classification-roberta-base>

²⁵<https://huggingface.co/CIRCL/vulnerability-severity-classification-chinese-macbert-base>

²⁶<https://huggingface.co/CIRCL/cwe-parent-vulnerability-classification-roberta-base>

With different CVSS scores

The image displays two browser windows side-by-side. The left window shows the CVE-2025-0108 entry on the [cvelistv5 - CVE-2025-0108](https://cvelistv5.cve.org/CVE-2025-0108) page. The right window shows the [Vulnerability Severity Classification](https://huggingface.co/spaces/CIRCL/vulnerability-severity-classification-RoBERTa-base) interface by CIRCL, which has classified the same vulnerability as 'High'.

Left Window: CVE-2025-0108 Details

CVE-2025-0108 (GCVE-0-2025-0108)
Vulnerability from [cvelistv5](#)

Published 2025-02-12 20:55
Modified 2025-07-30 01:36
Severity ? **8.8 (High)** - CVSS:4.0/AV:N/AC:L/AT:N/PR:N
5.9 (Medium) - CVSS:4.0/AV:N/AC:L/AT:P/PR:N
High (confidence: 0.7843)

VLA1 Severity ? 94.01% (0.99889)
EPSS score ? **CWE-306** - Missing Authentication for Critical Function
CWE **Summary** An authentication bypass in the Palo Alto Networks PAN-OS software enables an unauthenticated attacker with network access to the management web interface and invoke certain PHP scripts. While invoking these PHP scripts does not enable remote code execution, it can negatively impact integrity and confidentiality of PAN-OS. You can greatly reduce the risk of this issue by restricting access to the management web interface to only trusted internal IP addresses according to our recommended best practices deployment guidelines <https://live.paloaltonetworks.com/t5/community-blogs/tips-and-tricks-how-to-secure-the-management-access-to-the-management-web-interface-of-PAN-OS/ta-p/138443>. This

References
► **URL**
► [Palo Alto Networks](#)

Impacted products **Vendor**
► [Palo Alto Networks](#)

CISA Known exploited vulnerability
Data from the [Known Exploited Vulnerabilities Catalog](#)
Date added: 2025-02-18

Right Window: Vulnerability Severity Classification

Enter a vulnerability description, and the model will classify its severity level.

text

An authentication bypass in the Palo Alto Networks PAN-OS software enables an unauthenticated attacker with network access to the management web interface to bypass the authentication otherwise required by the PAN-OS management web interface and invoke certain PHP scripts. While invoking these PHP scripts does not enable remote code execution, it can negatively impact integrity and confidentiality of PAN-OS. You can greatly reduce the risk of this issue by restricting access to the management web interface to only trusted internal IP addresses according to our recommended best practices deployment guidelines <https://live.paloaltonetworks.com/t5/community-blogs/tips-and-tricks-how-to-secure-the-management-access-to-the-management-web-interface-of-PAN-OS/ta-p/138443>. This

Clear **Submit**

output

High

High 70%
Medium 28%
Low 1%
Critical 1%

Vulnerability in Vulnerability-Lookup — published 25 September 2025

The screenshot displays two browser windows. The left window shows the 'vulnerabilitylookup' website with details for CVE-2025-60249 (GCVE-0-2025-60249). The right window shows details for CVE-1-2025-0004 (GCVE-1-2025-0004).

CVE-2025-60249 (GCVE-0-2025-60249)
Vulnerability from [cvelists5](#)

Published: 2025-09-25 00:00
Modified: 2025-09-25 17:14
Severity: **6.4 (Medium)** - [CVSS:3.1/AV:N/AC:L/AT:N/PR:N/UI:N/VC:H/VI:H/HA:H/SC:H/SI:H/SA:H](#)
VLA1 Severity: **Medium (confidence: 0.9/45)**
EPSS score: Not yet available from FIRST.
CWE: [CWE-79](#) - Improper Neutralization of Scripting

Summary
vulnerabilitylookup 2.16.0 allows vulnerabilitylookup instance with XSS vulnerability was discovered in the Comments, Bundles, and Sightings components. Untrusted data was not properly sanitized before being rendered in templates and tables, which could allow attackers to inject arbitrary JavaScript into the application. The issue was due to unsafe use of innerHTML and insufficient validation of dynamic URLs and model fields. This vulnerability has been fixed by escaping untrusted data, replacing innerHTML assignments with safer DOM methods, encoding URLs with encodeURIComponent, and improving input validation in the affected models.

References
► [URL](#)

Impacted products

Vendor	Product	Version
CIRCL	vulnerabilitylookup	Version: View

[Show details on NVD website](#)

[JSON](#) [Share](#) [Add a sighting](#) [To clipboard](#) [Edit](#)

CVE-1-2025-0004
Vulnerability from [gna.1](#)

Published: 2025-09-25 14:10
Modified: 2025-09-25 14:10
Severity: **10.0 (Critical)** - [CVSS:4.0/AV:N/AC:L/AT:N/PR:N/UI:N/VC:H/VI:H/HA:H/SC:H/SI:H/SA:H](#)
VLA1 Severity: **Medium (confidence: 0.85/14)**
EPSS score: Not yet available from FIRST.
CWE: [CWE-79](#) - Improper Neutralization of Input During Web Page Generation (XSS or 'Cross-site Scripting')

Summary
A cross-site scripting (XSS) vulnerability was discovered in the handling of user-supplied input in the Comments, Bundles, and Sightings components. Untrusted data was not properly sanitized before being rendered in templates and tables, which could allow attackers to inject arbitrary JavaScript into the application. The issue was due to unsafe use of innerHTML and insufficient validation of dynamic URLs and model fields. This vulnerability has been fixed by escaping untrusted data, replacing innerHTML assignments with safer DOM methods, encoding URLs with encodeURIComponent, and improving input validation in the affected models.

References
► [URL](#)

Impacted products

Vendor	Product	Version
CIRCL	vulnerabilitylookup	Version: View

[JSON](#) [Share](#) [Add a sighting](#) [To clipboard](#) [Edit](#) [Delete](#)

Patched in Cambridge at Vuln4Cast 2025 in the afternoon.

<https://github.com/vulnerabilitylookup/vulnerabilitylookup/commit/afa12347f1461d9481eba75ac19897e80a9c7434>

Few information (reserved 17 September 2025 - sightings since 9 September)

CVE-2025-57623 (GCVE-0-2025-57623)
Vulnerability from [cvelistv5](#)

Published	2025-09-25 00:00
Modified	2025-09-25 17:28
Severity ?	
VLAIR Severity ?	High (confidence: 0.957)
EPSS score ?	Not yet available from FIRST.
CWE	n/a
Summary	A NULL pointer dereference in TOTOLINK N600R firmware v4.3.0cu.7866_B2022506 allows attackers to cause a Denial of Service.

References

- URL

Tags

Impacted products	Vendor	Product	Version
	n/a	n/a	Version: n/a

[Show details on NVD website](#)

JSON ▼ Share ▼ Add a sighting ▼ To clipboard Edit

Related vulnerabilities 0 Comments 0 Bundles 0 Sightings 2 Sighting correlations CWEs and mitigations

MITRE EMB3D

Sightings

Author	Source	Type	Date
automation	https://infosec.exchange/users/cr0w/statuses/115266254335547502 (correlations)	seen	55 minutes ago
automation	https://gist.github.com/z472421519/d17061ea79a72d39fe69c000fa1a6280 (correlations)	seen	16 days ago

Nomenclature

CWE Guessing (GCVE-1-2025-0004 - CWE 79)

CWE PARENT Patch Vulnerability Patch Classification Roberta Base - a Hugging Face Space by CIRCL — Mozilla Firefox

CWE PARENT Patch Vulner: x +

huggingface.co/spaces/CIRCL/CWE-PARENT-patch-Vulnerab

Spaces CIRCL/CWE-PARENT-patch-Vulnerability-Patch-Classification-Roberta-Base like 0

Running

CWE Prediction from Commit Message or Vulnerability Description

This tool uses a fine-tuned model to predict CWE categories from Git commit messages and vulnerability descriptions. Predicted child CWEs are mapped to their parent CWEs if applicable.

commit_message

A cross-site scripting (XSS) vulnerability was discovered in the handling of user-supplied input in the Comments, Bundles, and Sightings components. Untrusted data was not properly sanitized before being rendered in templates and tables, which could allow attackers to inject arbitrary JavaScript into the application. The issue was due to unsafe use of `innerHTML` and insufficient validation of dynamic URLs and model fields. This vulnerability has been fixed by escaping untrusted data.

Clear Submit

output

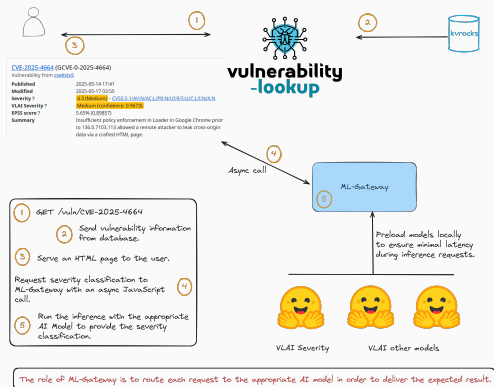
CWE-74

CWE-74	42%
CWE-436	8%
CWE-913	6%
CWE-664	6%
CWE-697	3%

Share via Link

Examples

A vulnerability has been found in cfire24 ajaxlife up to 0.3...



- Optional integration
- No dependencies with Vulnerability-Lookup
- Models are pulled from Hugging Face and preloaded locally
- Documented API (OpenAPI) to trigger the inferences
- <https://github.com/vulnerability-lookup/ML-Gateway>

Example

```
$ curl -X 'POST' \  
  'https://vulnerability.circl.lu/api/vlai/severity-classification' \  
-H 'accept: application/json' \  
-H 'Content-Type: application/json' \  
-d '{  
  "description": "An authentication bypass in the API component of Ivanti Endpoint  
    Manager Mobile 12.5.0.0 and prior allows attackers to access protected  
    resources without proper credentials via the API."  
}'  
{"severity": "High", "confidence": 0.8225}
```

Demo.

Closing

Future Development

- Deeper analysis of the content and context surrounding sightings and exploited vulnerabilities.
- Generation of insights based on patches.
- CPE Guessing.
- Allocation of vulnerability identifiers aligned with the GCVE system.
- Full-text search capabilities across all sources.
- Synchronization between multiple Vulnerability-Lookup instances.



The project is evolving rapidly — we always welcome feedback and feature suggestions!

References

 <https://www.vulnerability-lookup.org>

 <https://vulnerability.circl.lu>

 <https://github.com/vulnerability-lookup/vulnerability-lookup>

 <https://social.circl.lu/@circl>

Thank you for your attention

- Issues, new sources of advisories or ideas:
`https://github.com/vulnerability-lookup/vulnerability-lookup`
- For support and questions, contact: `info@circl.lu`