CIRCL
Computer Incident
Response Center
Luxembourg

GCVE.eu

Beyond CVEs: Mastering the Landscape with Vulnerability-Lookup

from CVE to CVD

https://www.vulnerability-lookup.org

Alexandre Dulaunoy - Cedric Bonhomme - team@circl.lu

October 23, 2025 - hack.lu 2025

CIRCL https://www.circl.lu 

TLP:CLEAR

Origin of the project

Who is behind Vulnerability-Lookup?

Vulnerability-Lookup[1] 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).
A reference implementation to **GCVE** standards.

vulnerability
-lookup

[1] https://www.vulnerability-lookup.org
[2] https://github.com/ngsoti

Origin

- `cve-search`[3] 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 at https://cve.circl.lu 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.

[3] `https://github.com/cve-search/cve-search`

Initial Challenges

- **Volume of data:** Handling a substantial dataset and heavy network traffic, currently over 1,360,500 security advisories and more than 90,000 sightings[4].
- **Flexibility:** Balancing ongoing development with legacy issues while designing a future-proof architecture. It's complex and yes, sometimes chaotic[5].
- **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- **Fast lookup:** Rapidly correlating identifiers across **diverse sources**, including unpublished advisories.

[4] The first sighting on Exploit-DB dates back 26 years.
[5] We enjoy challenges, especially when they lead to practical solutions.

Ongoing Challenges and Development

- **CPE fragmentation:**[6] 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 that fully supports the Coordinated Vulnerability Disclosure (CVD) process.[7]
- **Vulnerability numbering:** Enabling a new distributed approach through the Global CVE Allocation System.[8]
- **Scoring vulnerabilities:** Aggregating a large volume of observations from diverse advisory types to improve vulnerability scoring.

[6] Well, another mess to clean up!
[7] Aligned with NIS 2 and the Cyber Resilience Act.
[8] `https://gcve.eu`

Current Sources in Vulnerability-Lookup

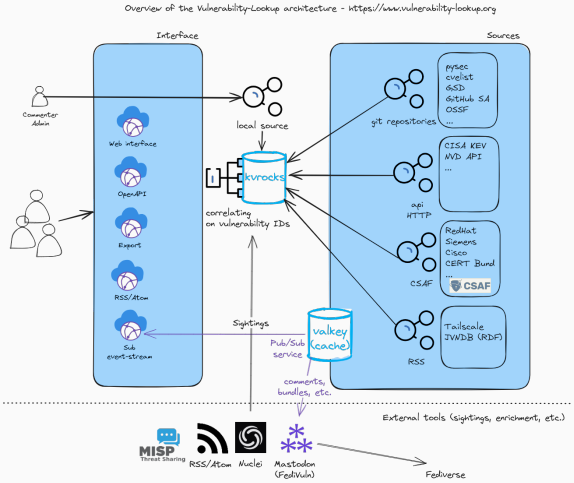
- **CISA Known Exploited Vulnerability** (HTTP)
- **NIST NVD CVE** (API 2.0)
- **CVEProject - cvelist** (Git submodule)
- **Fraunhofer FKIE** (Git submodule)
- **Cloud Security Alliance - GSD** (Git submodule)
- **GitHub Advisory DB** (Git submodule)
- **PySec Advisory DB** (Git submodule)
- **CSAF 2.0** (HTTP CSAF)
  CERT-Bund, Cisco, Siemens, Red Hat, Microsoft, NCSC-NL, CISA, etc.
- **VARIoT** (API)
- **Japan - JVN DB** (HTTP)
- **Tailscale** (RSS)
- **GCVE.eu all GNA sources**
- **CWE, CAPEC, MITRE EMB3D or KEV**
- **Growing...**

Open Data Initiative: Regular JSON dumps published[9].

[9] https://vulnerability.circl.lu/dumps/

Design and Implementation

Vulnerability-Lookup High-Level Architecture



TLP:CLEAR

Extended API

```
$ curl -s https://vulnerability.circl.lu/api/vulnerability/last/csaf_redhat/10 | jq .[2].document.title
"Red Hat Security Advisory: Red Hat Ceph Storage 6.1 security and bug fix update"

$ curl -s https://vulnerability.circl.lu/api/vulnerability/last/csaf_redhat/10 | jq .[2].vulnerabilities[0].cve
"CVE-2021-4231"
```

- **Documented API** (OpenAPI): https://vulnerability.circl.lu/api
- Pagination and filtering by source
- CPE search by vendor and product name
- **Many endpoints available via RSS and Atom**[10]

[10] https://www.vulnerability-lookup.org/documentation/feeds.html

Empowering the Community

Crowd-Sourced Threat Intelligence

- **Bundles:** Group similar vulnerabilities and aggregate sightings for easier tracking.
- **Comments:** Additional context such as PoCs, remediations, related insights.
- **Tags:** Use the MISP Vulnerability Taxonomy to annotate comments[11]. Example:

```
vulnerability:information=remediation
```

- **Sightings:** Report real-world observations of vulnerabilities, including metadata like timestamps and sources.


```
{
  "uuid": "f9ec8b2c-2ceb-4c05-b052-264b51c6a3ee", "vulnerability_lookup_origin": "1a89b78e-f703-45f3-bb86-59eb712668bd",
  "author": "9f56dd64-161d-43a6-b9c3-555944290a09", "creation_timestamp": "2025-04-17T19:14:32.000000Z",
  "vulnerability": "CVE-2025-32433",
  "type": "exploited",
  "source": "https://gist.github.com/numanturle/b7333fb02a4ee3618995babc9b62c507"
}
```


[11] https://www.misp-project.org/taxonomies.html#_vulnerability_3

Types of Sightings

Type	Description	Negative/Opposite
`seen`	The vulnerability was mentioned, discussed, or observed by the user.	-
`confirmed`	The vulnerability has been verified by an analyst.	X
`exploited`	The vulnerability was actively exploited and observed by the user reporting the sighting.	X
`patched`	The vulnerability was successfully mitigated or patched by the user reporting the sighting.	X

Table 1: Types of vulnerability sightings

Automated Sightings: Tools and Sources

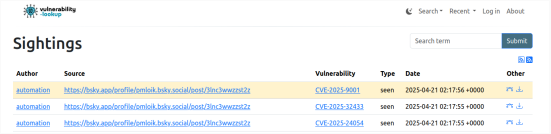
Automatically gathering crowd-sourced intelligence without requiring direct user contributions to our platform.

- **Social Platforms:** Fediverse, Bluesky
- **Threat Intelligence Tools:** MISP, Nuclei
- **Content Feeds:** RSS/Atom, curated web pages, GitHub Gist
- **Specialized Projects:** ShadowSight, ExploitDBSighting
- **Community Contributions:** Passive signals and indirect data enrichment

Scoring Vulnerabilities

Sightings Detection Rate and Types of Sightings

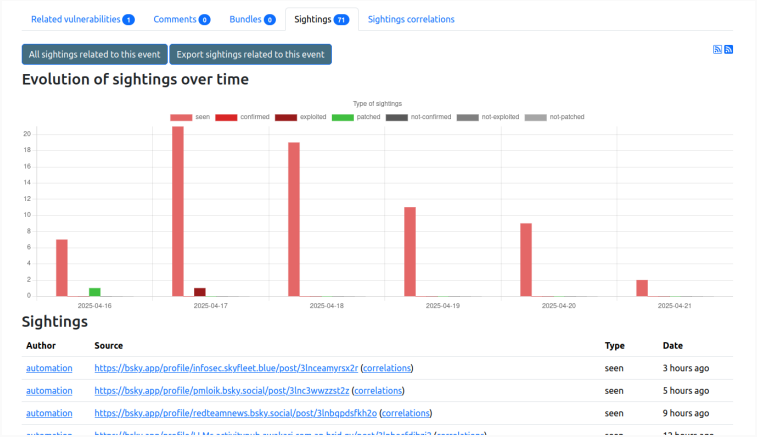
- A high rate of sightings (type *seen*) often correlates with high or critical severity vulnerabilities[12].
- Early sightings of type *exploited* (e.g., proof-of-concept code) or *confirmed* (e.g., detection templates for tools like Nuclei) can signal emerging threats.
- Sightings can sometimes be detected **before any official advisory is published**.



- Continuous exploitation patterns are frequently observed through sources like The Shadowserver Foundation or MISP.

[12] Don't underestimate the hype surrounding some vulnerabilities.

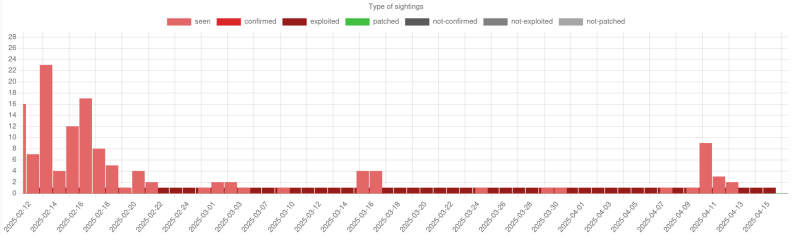
Early PoC (erlang / otp)



https://vulnerability.circl.lu/vuln/CVE-2025-32433#sightings

Continuous Exploitations (Palo Alto Networks / Cloud NGFW)

Evolution of sightings over time



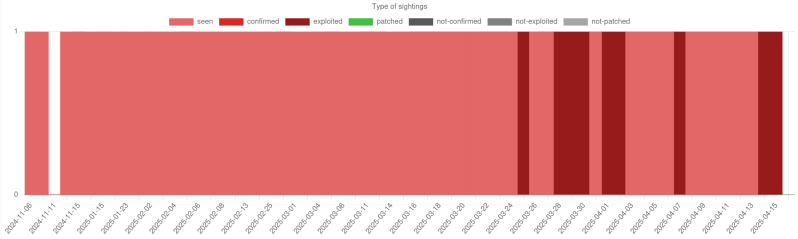
Sightings

Author	Source	Type	Date
automation	The Shadowserver (honeypot/exploited-vulnerabilities) - (2025-04-16) (correlations)	exploited	1 day ago
automation	https://bsky.app/profile/christopherkunz.bsky.social/post/3lmu2zatyx22z (correlations)	seen	2 days ago
automation	https://chaos.social/users/christopherkunz/statuses/114340622271163262 (correlations)	seen	2 days ago
automation	The Shadowserver (honeypot/exploited-vulnerabilities) - (2025-04-15) (correlations)	exploited	2 days ago

https://vulnerability.circl.lu/vuln/CVE-2025-0108#sightings

Continuous Exploitations (D-Link / DNS-320)

Evolution of sightings over time

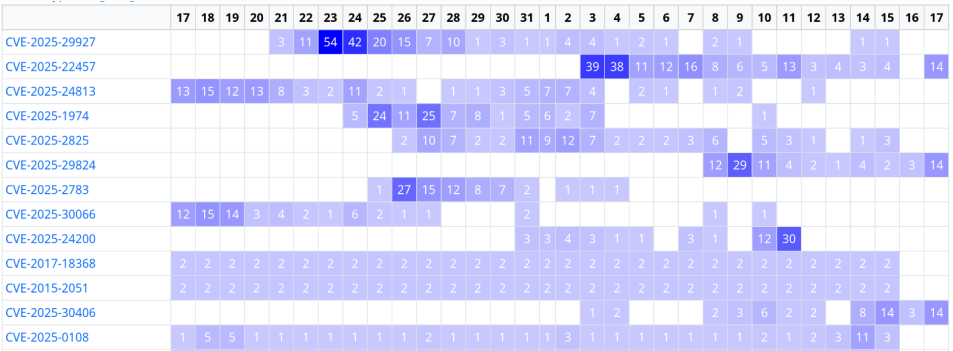


Sightings

Author	Source	Type	Date
automation	The Shadowserver (honeypot/exploited-vulnerabilities) - (2025-04-16) (correlations)	exploited	1 day ago
automation	The Shadowserver (honeypot/common-vulnerabilities) - (2025-04-16) (correlations)	seen	1 day ago
automation	The Shadowserver (honeypot/common-vulnerabilities) - (2025-04-15) (correlations)	seen	2 days ago
automation	The Shadowserver (honeypot/exploited-vulnerabilities) - (2025-04-15) (correlations)	exploited	2 days ago

https://vulnerability.circl.lu/vuln/CVE-2024-10914#sightings

Last Month's Most Sighted Vulnerabilities



	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CVE-2025-29927					3	11	54	42	20	15	7	10	1	3	1	1	4	4	1	2	1		2	1					1	1		
CVE-2025-22457																		39	38	11	12	16	8	6	5	13	3	4	3	4		14
CVE-2025-24813	13	15	12	13	8	3	2	11	2	1		1	1	3	5	7	7	4		2	1		1	2			1					
CVE-2025-1974								5	24	11	25	7	8	1	5	6	2	7							1							
CVE-2025-2825									2	10	7	2	2	11	9	12	7	2	2	2	3	6		5	3	1		1	3			
CVE-2025-29824																							12	29	11	4	2	1	4	2	3	14
CVE-2025-2783									1	27	15	12	8	7	2		1	1	1													
CVE-2025-30066	12	15	14	3	4	2	1	6	2	1	1				2								1		1							
CVE-2025-24200															3	3	4	3	1	1		3	1		12	30						
CVE-2017-18368	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
CVE-2015-2051	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
CVE-2025-30406																		1	2				2	3	6	2	2		8	14	3	14
CVE-2025-0108	1	5	5	1	1	1	1	1	1	1	2	1	1	1	1	1	3	1	1	1	1	1	1	1	2	1	2	3	11	3		

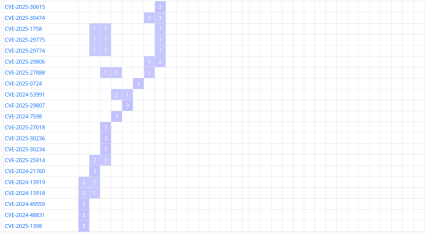
- **CVE-2025-22457:** Ivanti / Connect Secure — Severity: 10.0 (Critical)
- **CVE-2025-29927:** Vercel / Next.js — Severity: 9.1 (Critical)

Other Examples

Vulnerability	Product	Sighting count	EPSS	Severity
CVE-2025-29927	next.js	167	89.24% (0.99521)	9.1
CVE-2025-24813	Apache Tomcat	128	93.55% (0.99827)	9.2
CVE-2024-4577	PHP	190	94.38% (0.99961)	9.8
CVE-2025-0282	Connect Secure	243	90.87% (0.99618)	9.0
CVE-2024-55591	FortiOS	126	92.79% (0.99756)	9.8
CVE-2024-10914	D-Link DNS-320	81	93.73% (0.9985)	9.2
CVE-2020-21650	Myucms	57	2.48% (0.83998)	9.1

Table 2: Top vulnerabilities from our April 2025 report, based on sightings and scoring data.

Least Sighted Vulnerabilities in the Last Month



- **Low-sighting outliers offer valuable intel**, even if absent from EPSS or predictive models.
- Particularly relevant in low-noise sources (e.g., MISP, private Telegram channels).
- Often rated low/medium by CVSS and have low EPSS scores.
- Trend highlights EPSS's dependence on public threat intel feeds.

Tracking the Exploitability of Vulnerabilities Prior to Public Disclosure

- **Google / Android:** https://vulnerability.circl.lu/vuln/CVE-2024-43093#sightings
- **Speedify VPN (macOS):** https://vulnerability.circl.lu/vuln/CVE-2025-25364#sightings
- **SourceCodester:** https://vulnerability.circl.lu/vuln/CVE-2025-3821#sightings
  - Low visibility, no EPSS score, few sightings

Toward Practical AI Applications

From Data to Datasets

Contents

Why We Share Datasets

- **Open Data Initiative**: CIRCL's commitment to making data openly available.
- Consistent open approach applied across all our projects.
- Regularly updated JSON dumps [13] and "AI" datasets [14].
- Public, unauthenticated API access for Vulnerability-Lookup.

[13] https://vulnerability.circl.lu/dumps/
[14] https://huggingface.co/CIRCL/datasets

The Messy Reality of Large Datasets

- Our experience with large datasets is not recent (Passive DNS[15], BGP ranking[16], MISP[17], AIL[18], Lookyloo[19], 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).

[15] https://www.circl.lu/services/passive-dns/
[16] https://github.com/D4-project/BGP-Ranking
[17] https://github.com/MISP
[18] https://github.com/ail-project
[19] https://github.com/Lookyloo

Building AI Datasets

- Turn messy data into structured, actionable insights.
- Link related vulnerabilities via enrichment, correlation, and crawling.
- Support the process with **VulnTrain**[20].

[20] https://github.com/vulnerability-lookup/VulnTrain

Current datasets

Dataset	Size (rows)	Generation Time	Features
vulnerability-scores[21]	641,547	10m45s	Descriptions (en), CVSS, CPE
vulnerability-CNVD[22]	122,546	1m31s	Descriptions (cn), CVSS
vulnerability-cwe-patch[23]	883	210m	Descriptions (en), CWE, patches (commit id + url + full diff)

[21] https://huggingface.co/datasets/CIRCL/vulnerability-scores
[22] https://huggingface.co/datasets/CIRCL/Vulnerability-CNVD
[23] https://huggingface.co/datasets/CIRCL/vulnerability-cwe-patch

From Datasets to Models

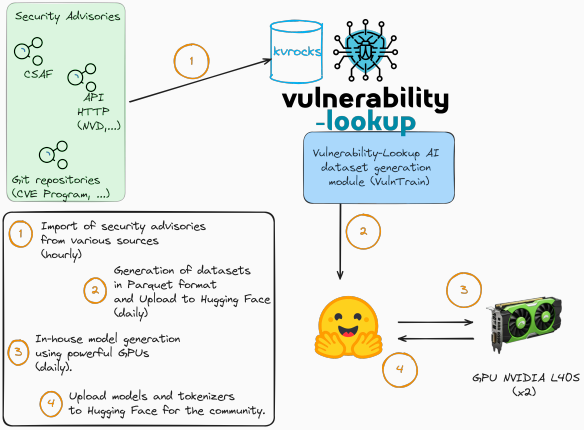
Contents

Why We Are Building AI Models

- CIRCL AI approach[24]: 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) project[25]

[24] https://circl.lu/pub/ai-strategy/
[25] https://www.science.nask.pl/en/research-areas/projects/12456

Model generation workflow



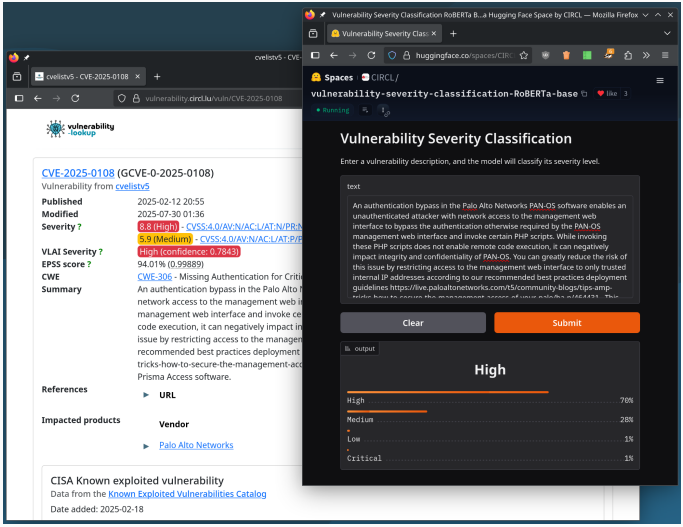
- local training
- models are publicly shared
- regular update

Current Models

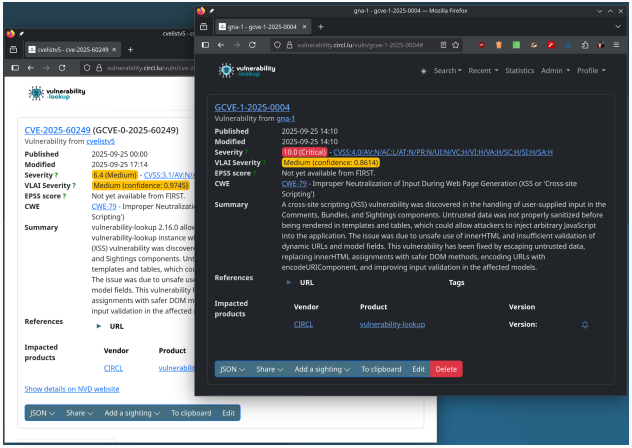
Model	Size	Epochs	Accuracy	Training Time
Severity classification[26]	125M params	5	0.8289	6.72h
Severity classification (CNVD)[27]	102M params	5	0.7817	65.989m
CWE guessing[28]	125M params	36-40	0.875	30m

[26] https://huggingface.co/CIRCL/vulnerability-severity-classification-roberta-base
[27] https://huggingface.co/CIRCL/vulnerability-severity-classification-chinese-macbert-base
[28] https://huggingface.co/CIRCL/cwe-parent-vulnerability-classification-roberta-base

With different CVSS scores



Vulnerability in Vulnerability-Lookup — published 25 September 2025

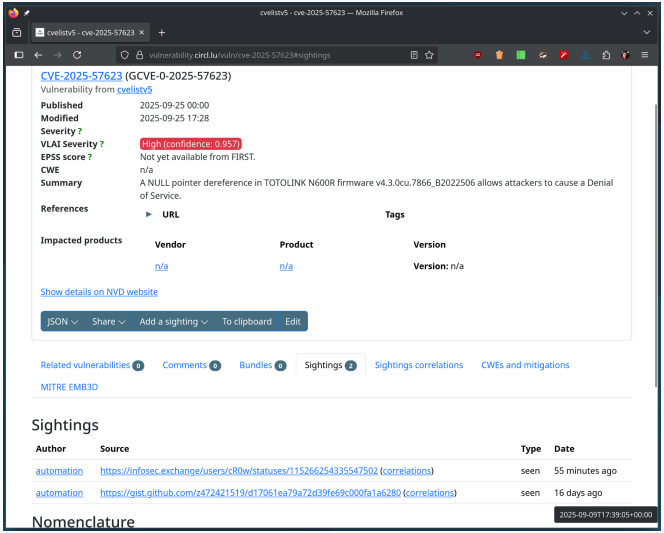


Patched in Cambridge at Vuln4Cast 2025 in the afternoon.

https://github.com/vulnerability-lookup/vulnerability-lookup/commit/afa12347f1461d9481eba75ac19897e80a9c7434

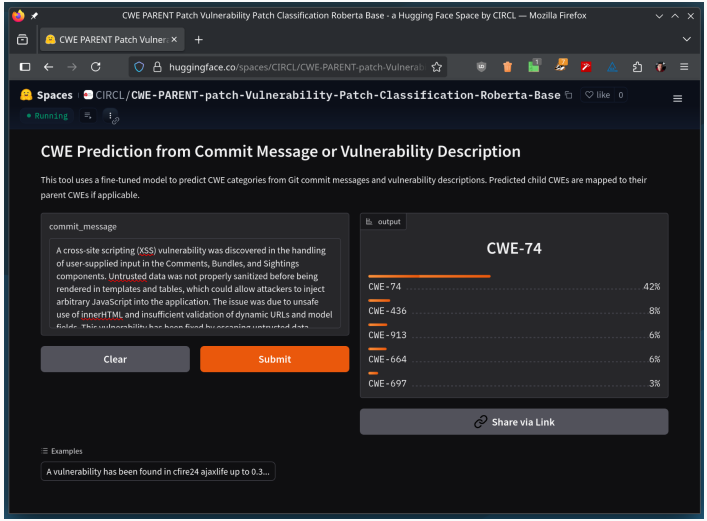
TLP:CLEAR

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

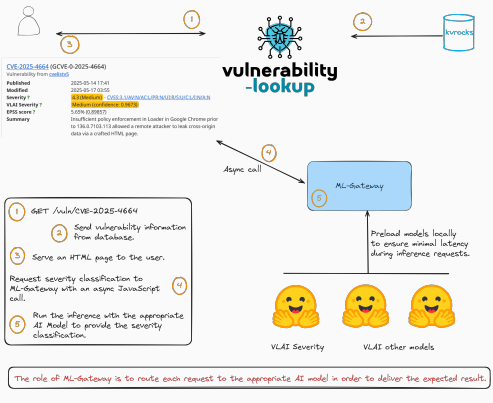


TLP:CLEAR

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



Integration



- 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}
```

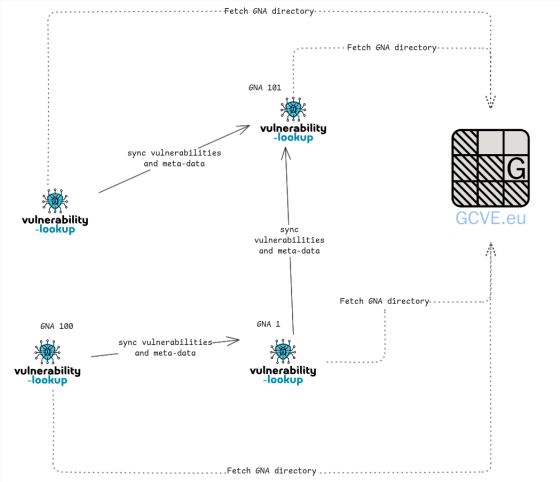
Lookup and AI are Cool, but Publishing is Even Cooler

GCVE.eu - Role

- The primary role of GCVE[29] is to provide **globally unique identifiers** to GCVE Numbering Authorities (GNAs).
- **GNAs operate autonomously**, with full control over how they assign and manage identifiers.
- **GCVE publishes Best Current Practices (BCPs)** on directory management, Coordinated Vulnerability Disclosure (CVD), and publication protocols.
- GCVE maintains and publishes the **official directory of all GNAs**, including their publication endpoints.

[29] `https://gcve.eu/`

Decentralized Publication Standard



Closing

Future Development

- Deeper analysis of the content and context of sightings, including **source reliability assessment**.
- Full-text search capabilities across all integrated sources.
- Integration of scoring models such as Vuln4Cast[30], with testing planned on our dataset to enhance reproducibility.
- **Improved notification capabilities** for newly observed vulnerabilities via webhooks.

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

[30] https://github.com/FIRSTdotorg/Vuln4Cast

References

https://www.vulnerability-lookup.org

CIRCL public instance https://vulnerability.circl.lu

Source code https://github.com/vulnerability-lookup/vulnerability-lookup

Dataset, AI Model Training, Models
https://github.com/vulnerability-lookup/VulnTrain