

AI at CIRCL

Building our own models for vulnerability intelligence

<https://circl.lu/pub/ai-strategy>

Cédric Bonhomme - cedric.bonhomme@circl.lu

April 28, 2026

Presented at International Committee of the Red Cross — Luxembourg

1. How CIRCL approaches AI
2. Building our own models
3. VLAI in action

How CIRCL approaches AI



We **enhance** existing tools rather than replace them with a chatbot.

- **Pragmatic:** AI is one tool among many, used where it actually helps analysts.
- **Open:** open data, open datasets, open models, open code.
- **Sovereign:** we train and host our models with our own hardware.
- **Small and specialised:** task-specific models, not general-purpose LLMs.
- **Strategy:** <https://circl.lu/pub/ai-strategy>

Vulnerability-Lookup, in 30 seconds

- Open Source project led by **CIRCL**, co-funded by the **European Union**.
- One place to **search**, **triage** and **track** vulnerabilities from 70+ sources.
- More than **1.7M security advisories** and **400k+ sightings** collected.
- Used by CSIRTs, SOCs, ENISA (EUVD), and many organisations.
- Reference implementation of the **GCVE** standards^a.

^a<https://gcve.eu>



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



Building our own models

From raw data to AI datasets

- Years of experience handling large security datasets (MISP, AIL, Lookyloo, Passive DNS, BGP Ranking. . .).
- We turn messy, multi-source vulnerability data into **structured, actionable** datasets.
- Datasets are **publicly shared** on Hugging Face¹.

Dataset	Rows	Features
vulnerability-scores	694,871	Descriptions (en), CVSS, CPE
vulnerability-CNVD	127,854	Descriptions (zh), CVSS
Vulnerability-FSTEC	82,960	Descriptions (ru), CVSS
vulnerability-cwe-patch	39,260	Descriptions, CWE, full patch diffs

¹<https://huggingface.co/CIRCL>

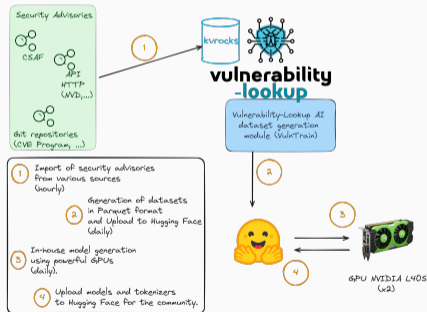
We train our own models



Custom models, trained **locally** on our own datasets.

- Stack: **PyTorch** + Hugging Face **Transformers**.
- Base models: **RoBERTa**, **MacBERT** (Chinese), **ruRoberta** (Russian)
- Fine-tuned on CIRCL datasets.
- Trained, versioned and published with **VulnTrain**^a.
- Paper: <https://arxiv.org/abs/2507.03607>

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



Models we have trained so far

Model	Params	Epochs	Accuracy	Train time
Severity classification (EN) ¹	125M	5	0.829	< 4h
Severity classification (ZH) ²	102M	5	0.782	< 1h
CWE guessing ³	125M	36–40	0.875	< 1h

All models are published openly — anyone can pull them, audit them, fine-tune them.

¹<https://huggingface.co/CIRCL/vulnerability-severity-classification-roberta-base>

²<https://huggingface.co/CIRCL/vulnerability-severity-classification-chinese-macbert-base>

³<https://huggingface.co/CIRCL/cwe-parent-vulnerability-classification-roberta-base>

VLAI in action

Severity prediction — example

The image shows two browser windows. The left window displays the 'vulnerability-lookup' page for CVE-2025-0108. The right window shows a 'Vulnerability Severity Classification' tool interface.

Vulnerability-lookup Data:

- CVE-2025-0108** (GCVE-0-2025-0108)
- 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/SC:N/IA:C/MA:C/CR:P/RS:N/UC:R/US:W/VA:N/VI:N/VS:N/WE:N/XX:N/ZE:N/ZZ:N
- VLAI Severity ?** High (confidence: 0.7843)
- EPSS score ?** 94.01% (0.99889)
- CWE** CWE-306 - Missing Authentication for Critical Function
- Summary** 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://www.paloaltonetworks.com/its/community/blog/ups-amp-secops-how-to-secure-the-management-access-to-pan-os>
- References**
 - ▶ **URL**
- Impacted products**
 - Vendor**
 - ▶ [Palo Alto Networks](#)
- CISA Known exploited vulnerability**
Data from the [Known Exploited Vulnerabilities Catalog](#)
Date added: 2025-02-18

Vulnerability Severity Classification Tool:

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

Input: 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://www.paloaltonetworks.com/its/community/blog/ups-amp-secops-how-to-secure-the-management-access-to-pan-os>

Output: High

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

Predicted severity and confidence directly inside Vulnerability-Lookup.

Useful even when data is scarce

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

Published 2025-09-25 00:00
Modified 2025-09-25 17:28

Severity ?
VLAJ 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](#) [Comments](#) [Bundles](#) [Sightings](#) [Sightings correlations](#) [CWEs and mitigations](#)

MITRE EMB3D

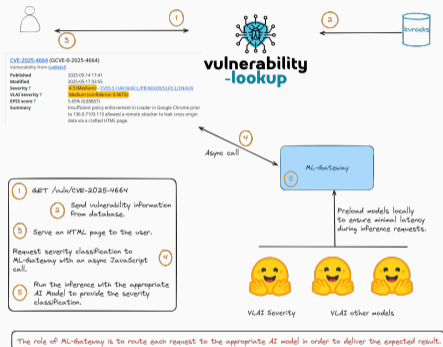
Sightings

Author	Source	Type	Date
automation	https://infosec.exchange/users/cr0w/statuses/115266254335542502 (correlations)	seen	55 minutes ago
automation	https://gist.github.com/2472421519/d17061ea79a72d39fe69c000fa1a6280 (correlations)	seen	16 days ago

Nomenclature 2025-09-09T17:39:05+00:00

Reserved CVE, sightings only — the model still gives a useful signal.

Integration via the ML-Gateway



- **Optional** component, decoupled from Vulnerability-Lookup.
- Models pulled from Hugging Face, preloaded locally.
- Documented **OpenAPI** for inference.
- Self-hosted: **your data stays with you.**
- <https://github.com/vulnerability-lookup/ML-Gateway>

CVE-2025-44897 (CVE-0-2025-44897)

Vulnerability from [cve.org](#)

Published 2025-05-20 00:00
Modified 2025-05-20 20:27

Severity ?

VLLM Severity ? **OpenJDK/JDK-8.0.920**

Summary

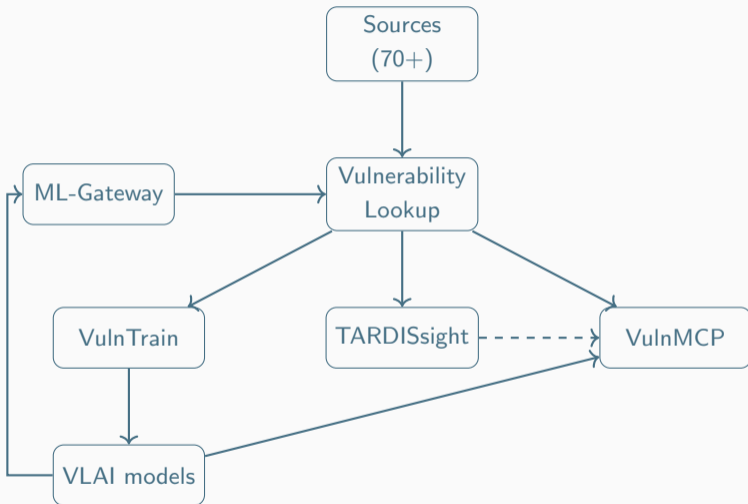
FW-WGS-804HPT v1.3056241111 was discovered to contain a stack overflow via the `lyftfp_srvip` parameter in the `web_tool_upgradeManager_posst` function.

The case of missing severity information in the advisory.

One HTTP call, one prediction

```
$ curl -X POST \  
  'https://vulnerability.circl.lu/api/vlai/severity-classification' \  
-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."  
}'  
  
{"severity": "High", "confidence": 0.8225}
```

From Vulnerability Data to Intelligence



Closing

What's next

- Deeper analysis of **sightings** and exploitation context.
- Insights generation from **patches** (commit + diff).
- **CPE guessing**
- **Assisted vulnerability description redacting**
- Continued participation in collaborative research (e.g. EU AIPITCH project).



Open data, open models, open code — feedback and contributions are welcome.

References

-  <https://www.vulnerability-lookup.org>
-  <https://vulnerability.circl.lu>
-  <https://github.com/vulnerability-lookup>
-  <https://huggingface.co/CIRCL>
-  <https://arxiv.org/abs/2507.03607>
-  <https://arxiv.org/abs/2604.16038>
-  GPU Efficiency in VLLM Model Training
-  <https://social.circl.lu/@circl>

Thank you

- Questions, ideas, new sources:
<https://github.com/vulnerability-lookup/vulnerability-lookup>
- Contact: info@circl.lu



CIRCL
Computer Incident
Response Center
Luxembourg