Integrating Autonomous Vulnerability Discovery into Open-Source Security Workflows

Stavros Kousidis

Master thesis · Published: 2026-10-15

Research question

How can findings from Cyber Reasoning Systems be validated and integrated into open-source vulnerability management and coordinated disclosure workflows?

Description

A Cyber Reasoning System may identify a suspected vulnerability or propose a patch. An automated report, however, is not yet evidence that the finding is exploitable, relevant to a maintained release, or sufficiently reliable for a maintainer to act upon. Open-source projects also have differing security policies and reporting workflows.

You will investigate how machine-generated findings can be independently checked, documented and prepared for human review. Particular attention should be paid to reproducibility, evidence quality, false-positive handling, patch regression testing and the separation between automated analysis and disclosure decisions.

Scope and evaluation

Select an accessible CRS or analysis pipeline and a small set of historical vulnerabilities in open-source projects. Work with locally isolated snapshots and compare automatically prepared reports against a defined reporting-quality rubric or baseline. Evaluate the effort needed for validation, the information useful to maintainers, and the cases in which automated evidence is insufficient.

The thesis must respect project security policies and the BSI's Coordinated Vulnerability Disclosure guidance. Testing current third-party services, contacting maintainers about new findings, or publishing exploit details is outside the default scope and requires prior approval and appropriate coordination.

Publication note: Follow upstream licenses and contribution procedures for proposed patches. Release independent evaluation tools under an agreed open-source license, but coordinate potentially sensitive vulnerability disclosures before publishing affected details. Open-source publication does not override responsible disclosure.

Expected outcome

An evaluated validation-and-reporting workflow, a reproducible collection of historical case studies, and recommendations for responsibly integrating CRS findings into open-source maintenance.

Starting points

- BSI: Coordinated Vulnerability Disclosure.
- AI Cyber Challenge open-source archive.
- Chin et al., *OSS-CRS: Liberating AIxCC Cyber Reasoning Systems for Real-World Open-Source Security* (2026).

Contact: stavros.kousidis@th-koeln.de
Code publication: open source by default; exceptions require prior agreement (policy).