ФГБОУ ВО «Алтайский государственный университет»

Разработка ML-инструмента для аннотирования патчей CVE

Выполнил: студент гр.4.207-1
Тихомирова З.В.

Научный руководитель: к.ф.-м. н., доцент Козлов Д.Ю.

Барнаул, 2026

Введение: актуальность и предметная область

(1) Common Vulnerabilities and Exposures (V) — единая система учёта уязвимостей

(2) CVE = запись об уязвимости

- **360 000+** CVE в NVD (июнь 2026)
- **34 000+** новых CVE в год
- CNA обязаны опубликовать запись в течение 24 часов

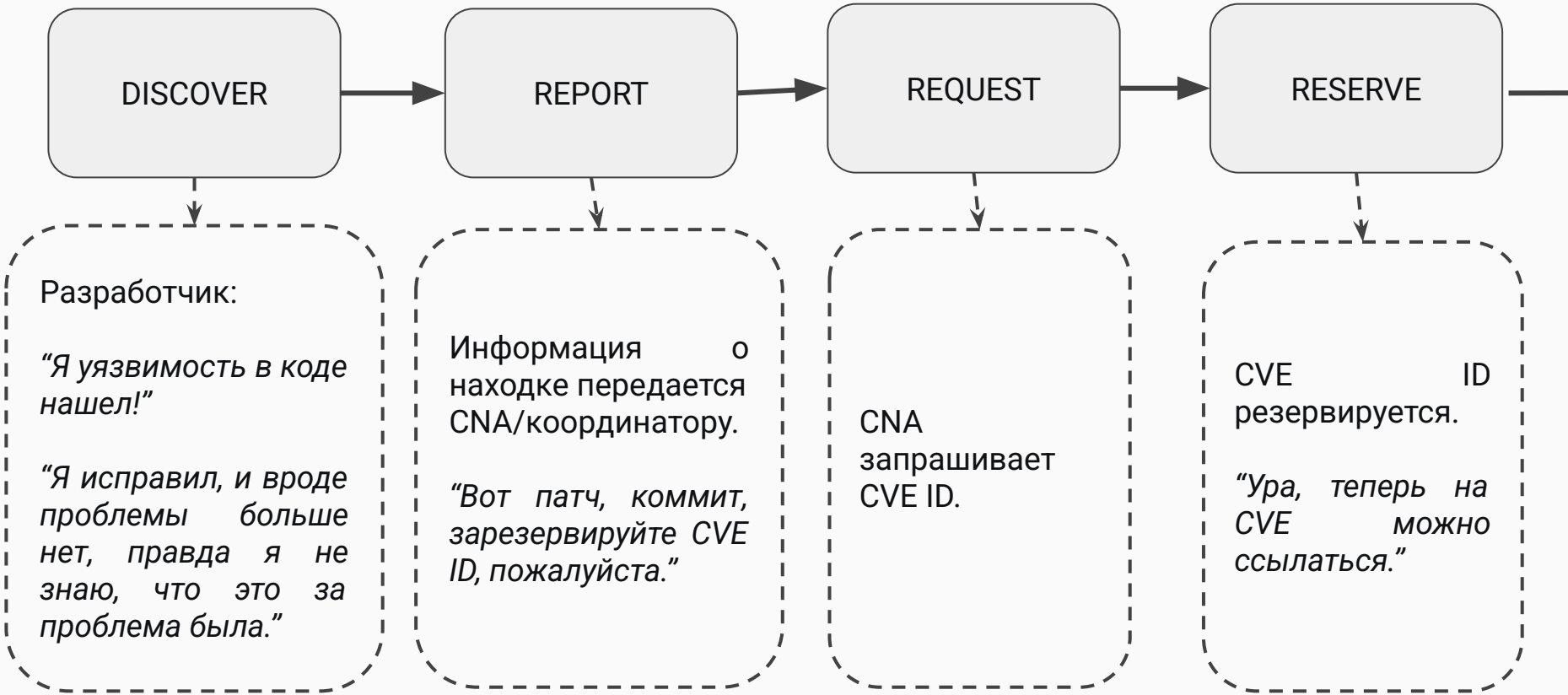
PUBLISHED

Year	2026	2025	2024	2023
Qtr4	TBA	12,796	11,073	7,876
Qtr3	TBA	11,738	8,591	6,936
Qtr2	TBA	11,701	11,716	7,134
Qtr1	15,176	12,009	8,697	7,015
TOTAL	15,176	48,244	40,077	28,961

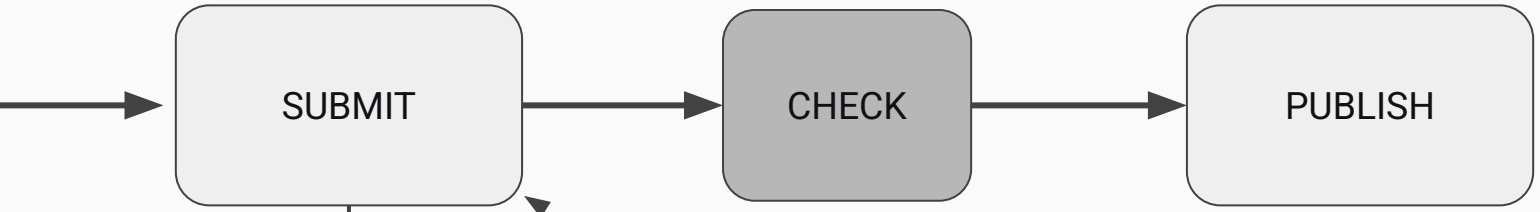
RECEIVED

Year	2026	2025	2024	2023
Count	21,530	70,729	52,316	40,051

Введение: актуальность и предметная область



Введение: актуальность и предметная область



CVE заполнена, можно отправлять на проверку.

!!! **BOTTLENECK** !!!

Ручной анализ одного патча — **часы**, сложного — **дни**.

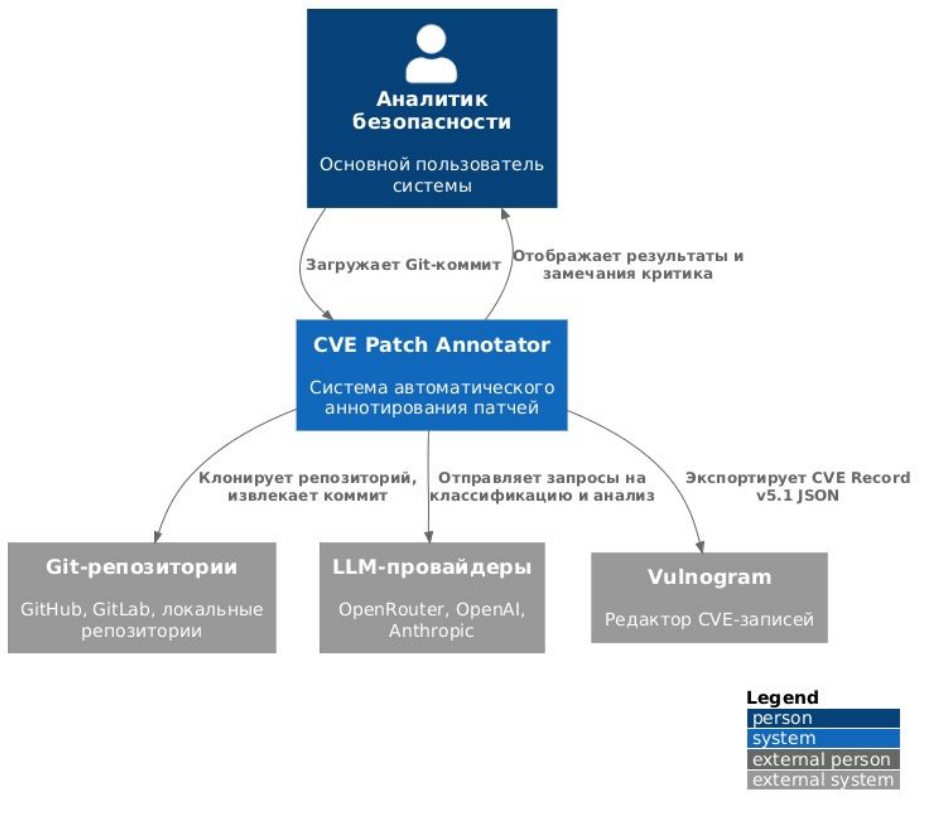
Автоматизированных **инструментов** для полного цикла **нет**.

CVE Patch Annotator

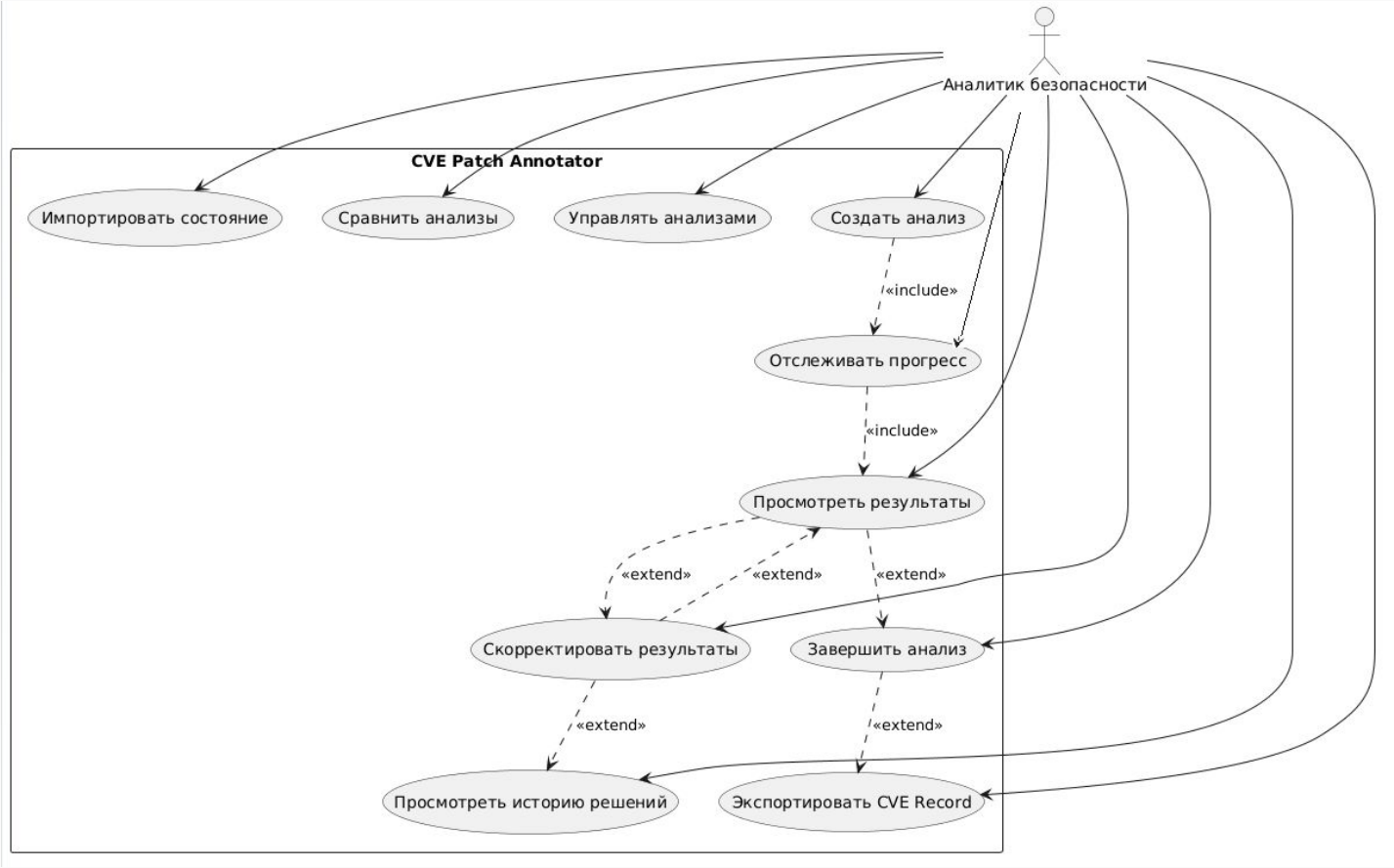
Разрабатываемая система — CVE Patch Annotator

Система решает проблему ручного аннотирования патчей безопасности с помощью мультиагентного подхода (12 агентов: 6 LLM-based, 6 детерминированных).

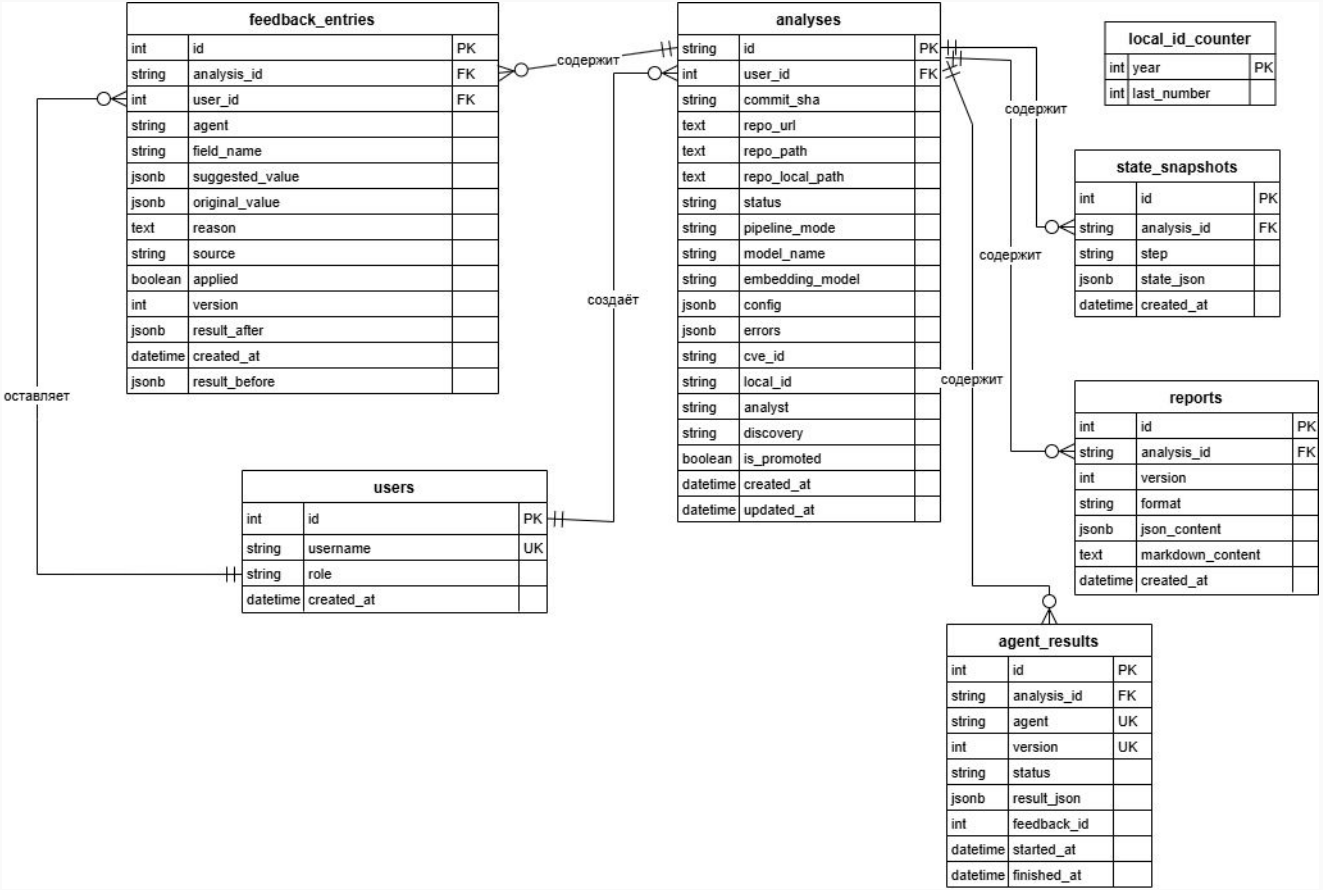
Справа — контекстная диаграмма системы.



Разрабатываемая система — CVE Patch Annotator. Use-Case



Разрабатываемая система — CVE Patch Annotator. ER-диаграмма



Разрабатываемая система — CVE Patch Annotator. Агенты

Детерминированные (6):

- A0 Git Provider — клонирует репозиторий, извлекает diff
- A1 Diff Parser — определяет языки, изменённые строки
- A2 AST Parser — извлекает изменённые функции (13 языков)
- A3 Context Expander — находит внешние зависимости
- A10 Version Tracer — определяет уязвимые версии
- A12 Report Builder — генерирует CVE Record v5.1

LLM-агенты (6):

- A4 Unit Analyzer — псевдокод SOURCE—SINK, taint-маркеры
- A5 Architecture Mapper — граф компонентов, Scope
- A6 CWE Classifier — гибридный поиск по 944 записям
- A7 CAPEC Classifier — выбор паттерна из 615 записей
- A8 CVSS Calculator — LLM-гипотеза + формула FIRST
- A11 Security Critic — кросс-валидация цепочки

Разрабатываемая система — CVE Patch Annotator. Агенты

PipelineOrchestrator:

1. 12 агентов выполняются последовательно.
2. Единый алгоритм для всех: проверка, сброс, выполнение, чекпоинт.
3. Атомарное сохранение после каждого агента.

Каскадная инвалидация (DOWNSTREAM_MAP):

1. Изменили CWE — CAPEC, CVSS, critic, report пересчитаны.
2. Изменили architecture — вся цепочка пересчитана.

Восстановление после сбоев:

1. Чекпоинт после каждого агента.
2. При перезапуске — продолжение с места остановки.

```
STEP_ORDER = [
    ("git", git.run_git),
    ("diff", diff.run_diff),
    ("ast", ast.run_ast),
    ("context", context.run_context),
    ("unit", unit.run_unit),
    ("architecture", architecture.run_architecture),
    ("cwe", cwe.run_cwe),
    ("capec", capec.run_capec),
    ("cvss", cvss.run_cvss),
    ("versions", versions.run_versions),
    ("critic", critic.run_critic),
    ("report", report.run_report),
]

MODEL_MAP = {
    "git": GitContext, "diff": DiffResult, "ast": ASTResult,
    "context": ContextResult, "unit": UnitAnalyzerOutput,
    "architecture": ArchitectureOutput, "cwe": CweResult,
    "capec": CapecOutput, "cvss": CvssOutput,
    "versions": VersionTracerOutput, "critic": SecurityCriticOutput,
    "report": ReportResult,
}

DOWNSTREAM_MAP = {
    "cwe":          ["capec", "cvss", "critic", "report"],
    "capec":        ["cvss", "critic", "report"],
    "cvss":         ["critic", "report"],
    "architecture": ["cwe", "capec", "cvss", "critic", "report"],
    "critic":       ["report"],
    "report":       [],
}
```

Разрабатываемая система — CVE Patch Annotator. Ключевые особенности

- ★ **Семантическое абстрагирование** — код превращается в псевдокод, downstream-агенты работают с семантикой, а не синтаксисом.
- ★ **Каскадная инвалидация** — изменил CWE — CAPEC, CVSS, critic и report пересчитались автоматически.
- ★ **Чекпоинты** — после каждого агента состояние сохраняется. Упал сервер — продолжили с места остановки.
- ★ **Состояние пайпплайна** — конструктор.
- ★ **Human-in-the-Loop** — два режима: override (запись значения) и retry (перезапуск агента с обратной связью). Аудит всех изменений.
- ★ **Развёртывание одной командой** — docker compose up -d.
- ★ **Гибридный поиск CWE** — три метрики одновременно: смысл (pgvector) + слова (tsvector) + пересечение текстов (Jaccard).

Разрабатываемая система — CVE Patch Annotator. Ключевые особенности

- ★ **CAPEC не случаен** — каждый кандидат получает вердикт MATCH/DISQUALIFIED. Все DISQUALIFIED, тогда генерируется CAPEC-GENERIC.
- ★ **CVSS**: две фазы — LLM-гипотеза + математическое ядро FIRST. Scope определяет Architecture Mapper, не LLM.
- ★ Создание отчетов в формате CVE Record v5.1.

Разрабатываемая система — CVE Patch Annotator. Технологический стек

Язык программирования и веб-фреймворк

FastAPI

База данных и поиск

Alembic

pgvector

SQLA

LLM и прокси

LangChain, OpenRouter API

LiteLLM

Синтаксический анализ

tree-sitter, unidiff

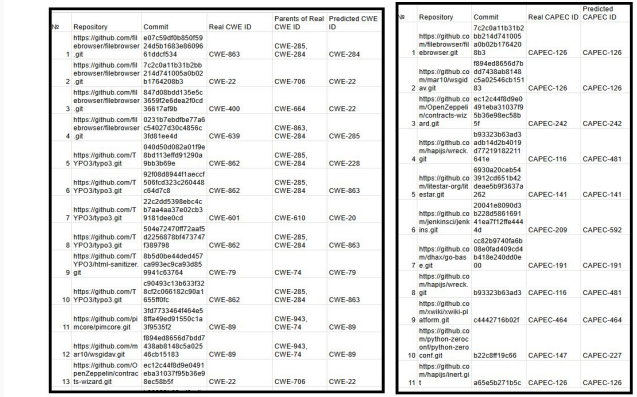
Веб-интерфейс

HTMX, Alpine.js, Jinja2, PicoCSS

Тестирование

Бенчмарк (30 записей GitHub Advisory Database):

- CWE: 60% точных (18 из 30), 63% с учётом близких категорий
- CAPEC: 63% (11 записей)
- CVSS (RMSE): 1.46 (13 записей)



	Repository	Commit	Real CWE ID	Parents of Real CWE ID	Predicted CWE ID
1	https://github.com/filebrowser/filebrowser.git	e07c59df0b850f592465b1683e86096616dc534	CWE-863	CWE-285, CWE-284	CWE-284
2	https://github.com/filebrowser/filebrowser.git	7c2c0a11b31b2bb214df41005ab02b176420b17e	CWE-22	CWE-706	CWE-22
3	https://github.com/filebrowser/filebrowser.git	847d08bdd135e5c365972e6dea2f0cd36617af9b	CWE-400	CWE-664	CWE-22
4	https://github.com/filebrowser/filebrowser.git	0231b7ebdfbe77a6c54027d30c4856c3fd81ee4d	CWE-639	CWE-863, CWE-284	CWE-285
5	https://github.com/TYPO3/typo3.git	0400d50d082a01f9e8bd113ef8d91290a9bb3b69e	CWE-862	CWE-285, CWE-284	CWE-228
6	https://github.com/TYPO3/typo3.git	9208d8944f1aeccf5986c6323c260448c64d7c8	CWE-862	CWE-285, CWE-284	CWE-863
7	https://github.com/TYPO3/typo3.git	22c2dd5398ebc4c7aa4aa37e02cb39181dee0cd	CWE-601	CWE-610	CWE-20
8	https://github.com/TYPO3/typo3.git	534e72470ff72aa5d2256878bf473747f389798	CWE-862	CWE-285, CWE-284	CWE-863
9	https://github.com/TYPO3/html-sanitizer.git	8b50dbe44ded457ca993ec9ca93d859941c63764	CWE-79	CWE-74	CWE-79
10	https://github.com/TYPO3/typo3.git	c90493c13b513f32 8cf2c066182c90a1655ff0fc	CWE-862	CWE-285, CWE-284	CWE-863
11	https://github.com/pimcore/pimcore.git	3fd7733464f464e58ffa49ed91550c1a3f9535f2	CWE-89	CWE-943, CWE-74	CWE-89
12	https://github.com/mar10/wsgidav.git	f884ed8656d7bdd7438ab8148c5a025 46cb15183	CWE-89	CWE-943, CWE-74	CWE-89
13	https://github.com/OpenZeppelin/contracts-wizard.git	ec12c44f8d9e0491eba31037f95b36e9 8ec58b5f	CWE-22	CWE-706	CWE-22

CWE

	Repository	Commit	Real CAPEC ID	Predicted CAPEC ID
1	https://github.com/filebrowser/filebrowser.git	7c2c0a11b31b2bb214df41005ab02b176420b17e8b3	CAPEC-126	CAPEC-126
2	https://github.com/mar10/wsgidav.git	f884ed8656d7bd7438ab8148c5a02546cb15183	CAPEC-126	CAPEC-126
3	https://github.com/OpenZeppelin/contracts-wizard.git	ec12c44f8d9e0491eba31037f95b36e98ec58b5f	CAPEC-242	CAPEC-242
4	https://github.com/hapijs/wreck.git	b93323b63ad3adb14d2b4019d77219182211641e	CAPEC-116	CAPEC-481
5	https://github.com/lifestar-org/lifestar.git	6830a20ceb543912cd651b42deae5b9f3637a262	CAPEC-141	CAPEC-141
6	https://github.com/jenkinsci/jenkins.git	20041e8090d3b228d5861691 41ea7f12ff8444 4d	CAPEC-209	CAPEC-592
7	https://github.com/dhax/go-base.git	cc82b97404fa6b08e0fad409cd4b418e240dd0e00	CAPEC-191	CAPEC-191
8	https://github.com/hapijs/wreck.git	b93323b63ad3	CAPEC-116	CAPEC-481
9	https://github.com/swiki/swiki-platform.git	c4442716b02f	CAPEC-464	CAPEC-464
10	https://github.com/python-zeronet/python-zeronet.git	b22c8ff19c66	CAPEC-147	CAPEC-227
11	https://github.com/hapijs/lnert.git	a65e5b271b5c	CAPEC-126	CAPEC-126

CAPEC

Анализ ошибок CVSS:

- Scope — определяется верно (A5 работает корректно)
- Attack Vector — систематически завышается
- Privileges Required — систематически занижается

Совместимость с Vulnogram протестирована.

Тестирование

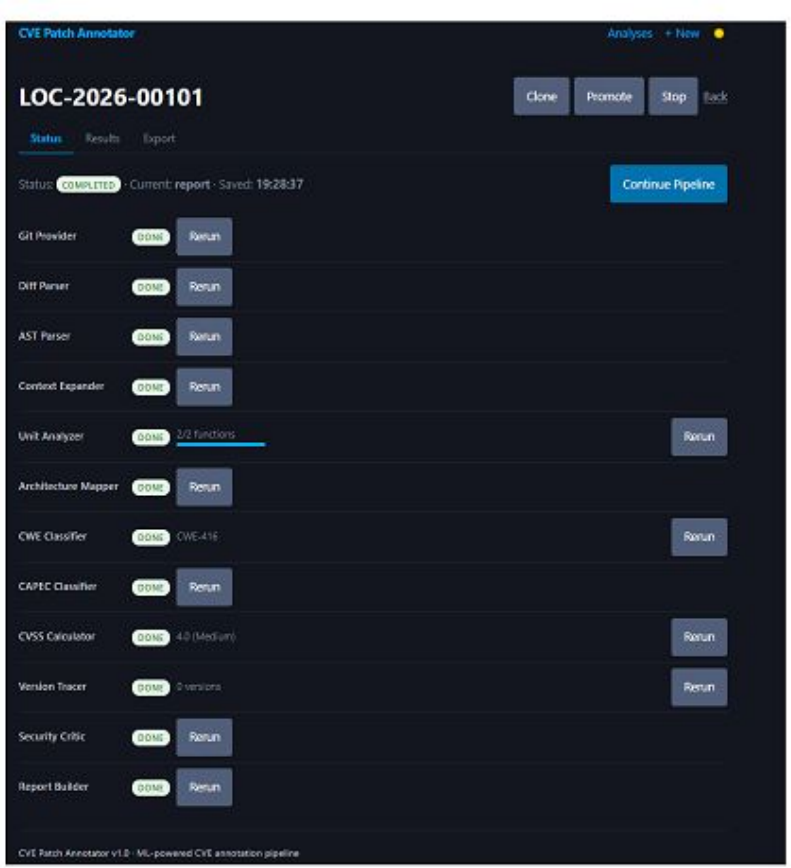
Модульное тестирование:

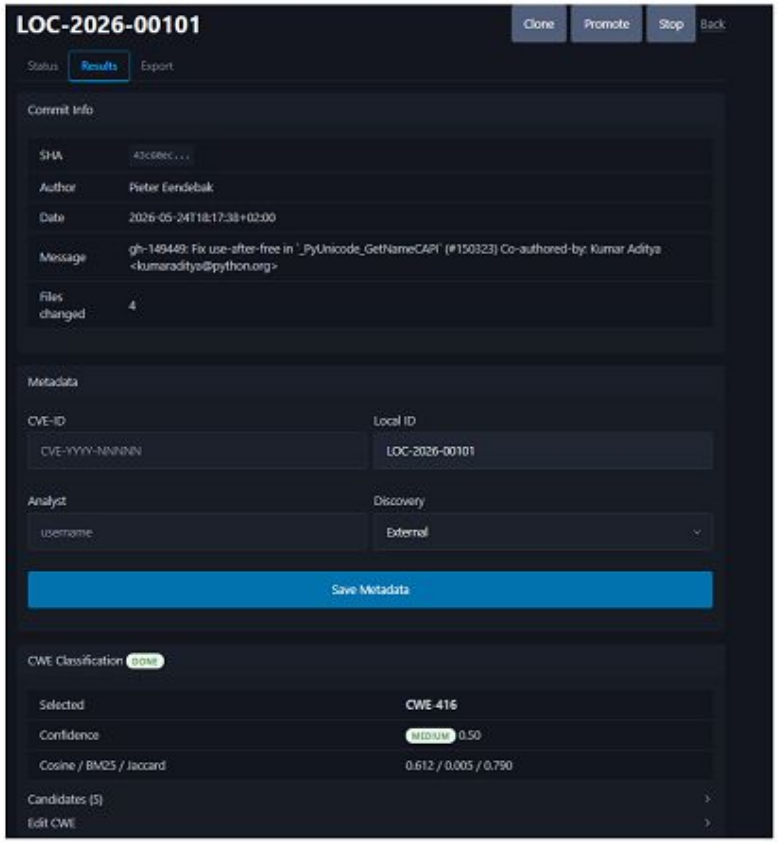
- 173 теста (pytest), покрытие кода 51%
  - детерминированные компоненты: Pydantic-схемы и модели — 100%, парсеры — до 82%
  - LLM-агенты: тесты отказоустойчивости (пустой вход, сбой LLM, rate limit)

Ручное системное тестирование:

- полный цикл через веб-интерфейс и CLI
- Загрузка коммита => 12 агентов => HITL-корректировка => экспорт CVE Record

Интерфейсы


CVE Patch Annotator
Analyses
+ New
LOC-2026-00101
Clone
Promote
Stop
Back
Status
Results
Export
Status: COMPLETED · Current: report · Saved: 19:28:37
Continue Pipeline
Git Provider
DONE
Rerun
Diff Parser
DONE
Rerun
AST Parser
DONE
Rerun
Context Expander
DONE
Rerun
Unit Analyzer
DONE
2/2 functions
Rerun
Architecture Mapper
DONE
Rerun
CWE Classifier
DONE
CWE-416
Rerun
CAPEC Classifier
DONE
Rerun
CVSS Calculator
DONE
4.0 (Medium)
Rerun
Version Tracer
DONE
2 versions
Rerun
Security Critic
DONE
Rerun
Report Builder
DONE
Rerun
CVE Patch Annotator v1.0 · ML-powered CVE annotation pipeline


LOC-2026-00101
Clone
Promote
Stop
Back
Status
Results
Export
Commit Info
SHA
42c66ec...
Author
Peter Eendebak
Date
2026-05-24T18:17:38+02:00
Message
gh-149449: Fix use-after-free in `_PyUnicode_GetNameCAPI` (#150323) Co-authored-by: Kumar Aditya <kumaraditya@python.org>
Files changed
4
Metadata
CVE-ID
CVE-YYYY-NNNNN
Local ID
LOC-2026-00101
Analyst
username
Discovery
External
Save Metadata
CWE Classification
DONE
Selected
CWE-416
Confidence
MEDIUM 0.50
Cosine / BM25 / Jaccard
0.612 / 0.005 / 0.790
Candidates (5)
Edit CWE

Заключение и выводы

Разработана система CVE Patch Annotator:

- 12 агентов: от git-коммита до CVE Record v5.1
- гибридный retrieval (pgvector + tsvector + Jaccard)
- Human-in-the-Loop с каскадной инвалидацией
- веб-интерфейс + CLI + Docker

Результаты:

- 173 теста, бенчмарк на 30 записях
- CVE Record v5.1 совместим с Vulnogram
- развёртывание одной командой

Система сокращает время рутинного анализа, оставляя эксперту содержательную работу.

Спасибо за внимание!

Интересные вопросы

1. Почему точность CWE 60%, а не 90%?
2. Почему самописный оркестратор, а не LangGraph?
3. Почему монолит, а не микросервисы?
4. Чем ваша система лучше Vulnogram?
5. Что будет, если LLM «придумает» несуществующий CWE?