Security Testing in the AI Era: The Smart Banka Application

Pavel Bartoň; MONETA Money Bank
Peter Koreň; AISLE

15. října 2026

ISACA
Czech Republic Chapter

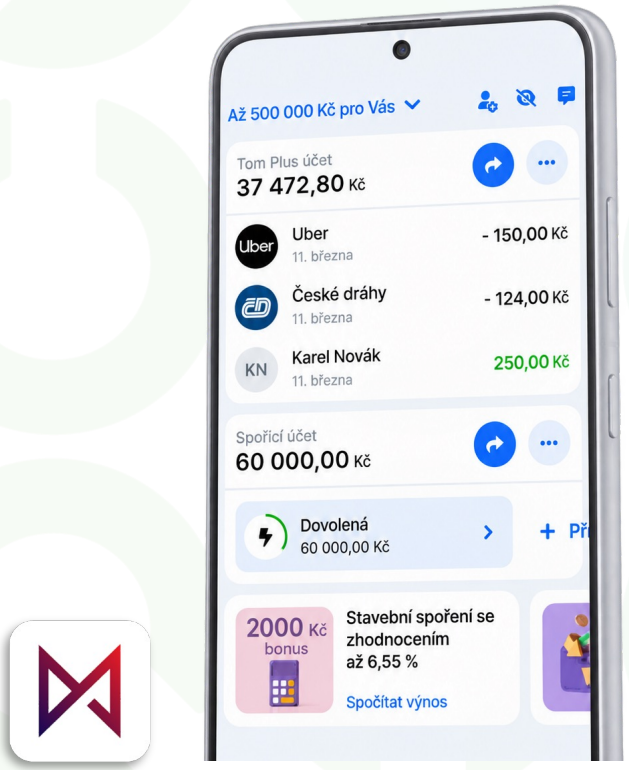
Smart Banka now serves more than 1.3 million registered users, making digital banking the primary customer channel at MONETA.

HOW DEVELOPMENT WORKS

- Frequent releases—the fastest cadence at MONETA
- New features are added almost daily
- One application supported by multiple teams

WHAT THIS MEANS FOR SECURITY

- A continuously evolving Kotlin-based front-end application
- More than 3,000 client-facing APIs, with new or modified endpoints
- New third-party integrations, SDKs and CI/CD capabilities
- Stricter regulation, including the Digital Operational Resilience Act (DORA) and Cyber Resilience Act (CRA)



ISACA Czech Republic Chapter

Continuous and AI-assisted testing extends the reach of traditional assessments.

TRADITIONAL PENETRATION TESTING

- Regular assessments by external specialists
- The full-scope approach was difficult to sustain
- After low-hanging fruit was fixed, more complex vulnerabilities remained

BUG-BOUNTY PROGRAMME

- Managed internally
- 500+ submissions (2023–2026)
- Revealed vulnerabilities that external specialists had not found
- Increasing number of submissions in AI era

WEEKLY FEATURE TESTING

- New features tested weekly
- Brought in-house because external delivery was too costly and inflexible

LLM-ASSISTED TESTING (OPEN-WEIGHT MODELS)

- Local LLMs analyzed obfuscated and de-obfuscated binaries and source code
- Developed attack vectors and improved threat models
- Continuous feedback loop
- Not Mythos-level (quite expensive for a bank in terms of hardware acquirement)

ISACA Czech Republic Chapter

With support from TRASK, we assessed the Smart Banka Android source code using AISLE's Mythos-class AI.

WHAT WE NEEDED

- Frontier-level LLM access beyond our existing open-weight models
- Regulatory and compliance alignment: zero data retention (ZDR) and EU-based tenants
- Not just access to the model but engineering help (context building, feedback)

WHY SMART BANKA FOR ANDROID

- Crown jewel of any bank
- Android's more open ecosystem can increase the attack surface compared with iOS
- We wanted to identify as many vulnerabilities as possible before bug-bounty researchers found them

ISACA Czech Republic Chapter


AISLE
ISACA
Czech Republic Chapter

Who is AISLE

AI-native AppSec platform: an autonomous system that finds, fixes and verifies vulnerabilities.

Goal: zero exploitable vulnerabilities.

FOUNDERS

Ondřej Vlček
CEO (CEO of Avast
President of Gen Digital)

Jaya Baloo
COO & CISO (Rapid7 CISO,
Top 100 CISOs)

Stanislav Fort
Chief Scientist (Anthropic,
DeepMind, Google Brain)

ANGEL INVESTORS & ADVISORS

Jeff Dean
Chief Scientist at Google

Thomas Wolf
Co-founder & Chief Scientist at Hugging Face

Olivier Pomel
CEO at Datadog

Aparna Chennapragada
CPO for AI Experiences at Microsoft

Ian Goodfellow
Research Scientist at DeepMind

Prof Surya Ganguli
Advisor · Stanford University AI Professor

Bruce Schneier
Advisor · Fellow at Harvard University

TRUSTED BY

OpenSSL Foundation
20 of 23 CVEs found in six months

curl
Code on 20B+ devices

ENISA
SRP platform and EUVD

Ethereum Foundation
Execution clients and consensus clients

ISACA Czech Republic Chapter

System > model

ISACA
Czech Republic Chapter

What AISLE does.

Find it. Fix it. Keep it fixed.

01

Discover

Reasons over all your code, so findings are real

02

Remediate

Drafts fixes, tests them, opens pull requests

03

Protect

Checks every new change before it merges

ISACA Czech Republic Chapter

Current State

ISACA
Czech Republic Chapter

2026 already beat the forecast.

MID-2026 FORECAST

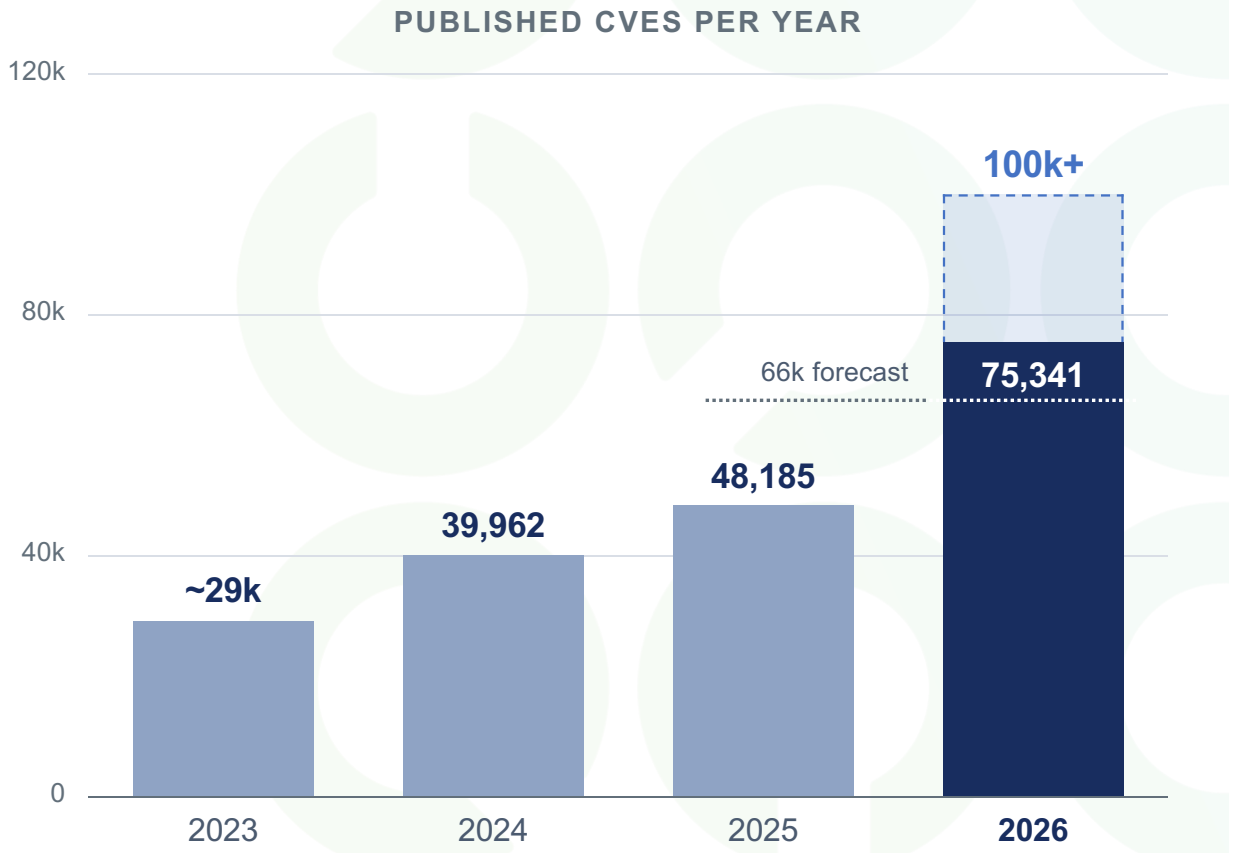
~~66,000~~

PUBLISHED SO FAR

75,341

NEW FORECAST

100,000+



Sources: Jerry Gamblin, FIRST, Trend Micro

ISACA Czech Republic Chapter

AI made attackers faster.

More vulnerabilities. Faster attackers.

ATTACK VOLUME

+89%

more attacks by AI-enabled adversaries

ZERO-DAY EXPLOITATION

42%

of exploited vulnerabilities were used before disclosure



TIME TO EXPLOIT

–7 days

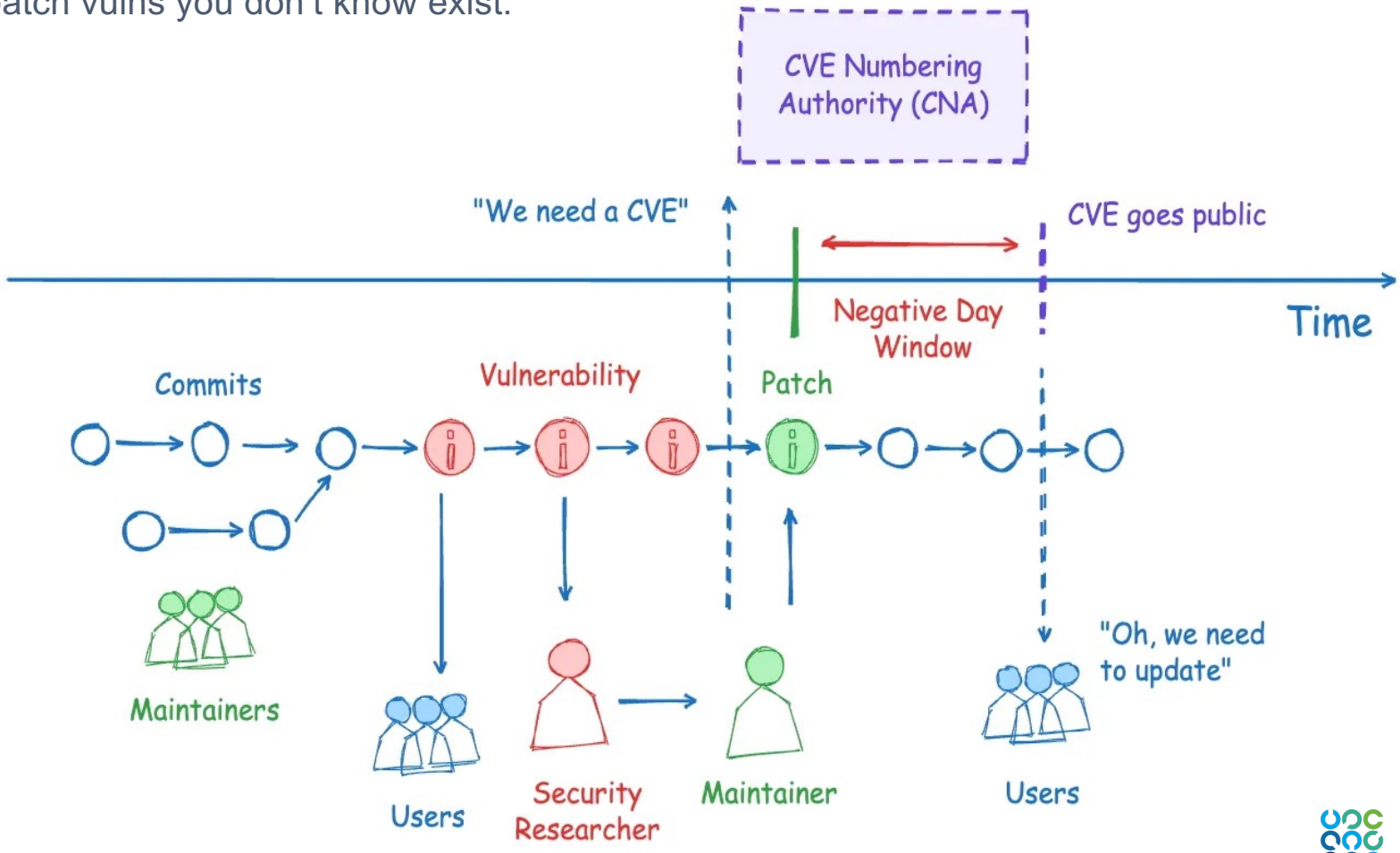
exploits are in the wild an average of 7 days before a patch exists

Sources: CrowdStrike 2026 Global Threat Report · Mandiant M-Trends 2026

ISACA Czech Republic Chapter

Negative-day windows and silent patches

You can’t patch vulns you don’t know exist.



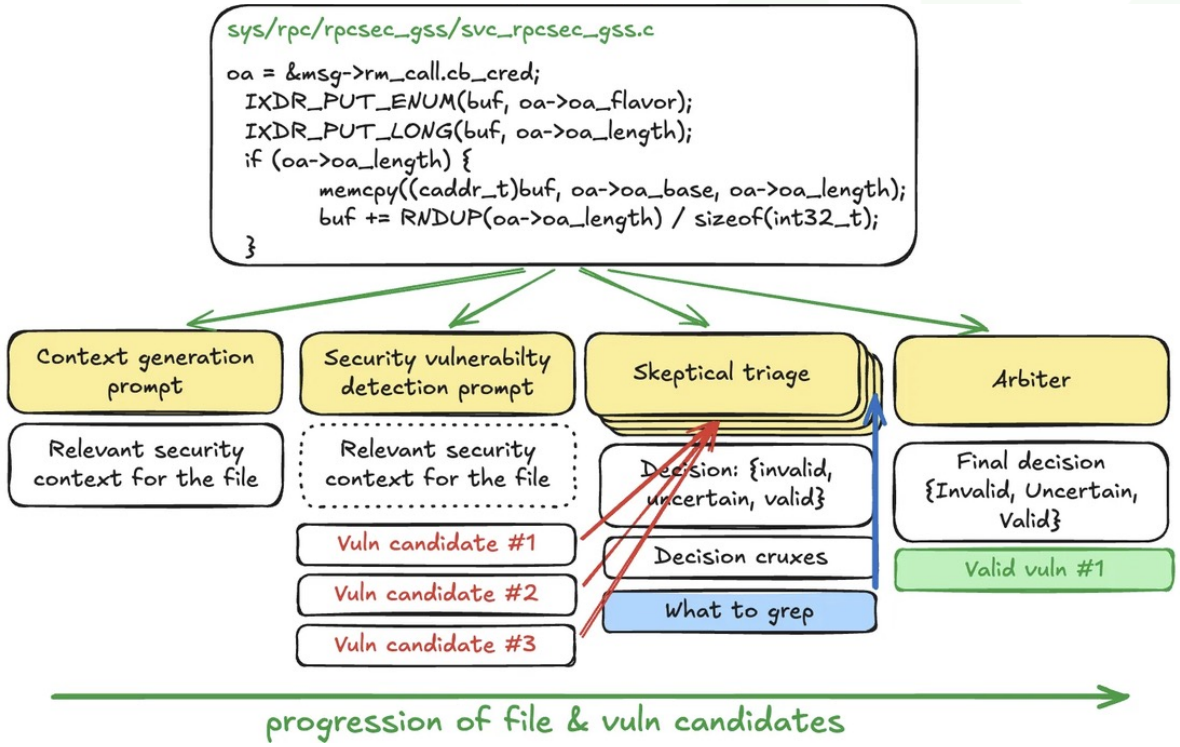
Source: aisle.com/blog/before-the-cve-mining-open-source-commits-for-negative-day-vulnerabilities

ISACA Czech Republic Chapter


Project Glasswing
Securing critical software for the AI era
Continue reading
ANTHROP\C
Claude Mythos
ISACA
Czech Republic Chapter

AISLE nano-analyzer

mini harness example: context, detection, triage and decision.



Source: github.com/weareaisle/nano-analyzer

ISACA Czech Republic Chapter

We reran Mythos's showcase bugs on small, open models.

Most found the same bugs.

FREEBSD RCE

8 / 8

models found Mythos's flagship bug in a single call

SMALLEST MODEL

3.6B

active parameters, at $0.11 per million tokens

OPENBSD SACK

5.1B

active-parameter open model recovered the full 27-year-old chain

> **Why:** with the right code in view, spotting a bug is a narrow task. It takes context more than raw intelligence.

Source: AISLE Research · The Jagged Frontier, Apr 2026

Prompting a model vs harness

PROMPTING AN LLM

- It finds some vulnerabilities. Are you sure, that it examined all the code?
- It does not learn new information over time
- Costs are not predictable and they change
- It does not help you manage findings

SYSTEM / HARNESS

Orchestration makes sure that the scan is complete
It examines all the code efficiently

Controlled and safe access
Approved and audited tools operate in a sandbox

A combination of different models
Models from different sources find different problems

Control and reliability

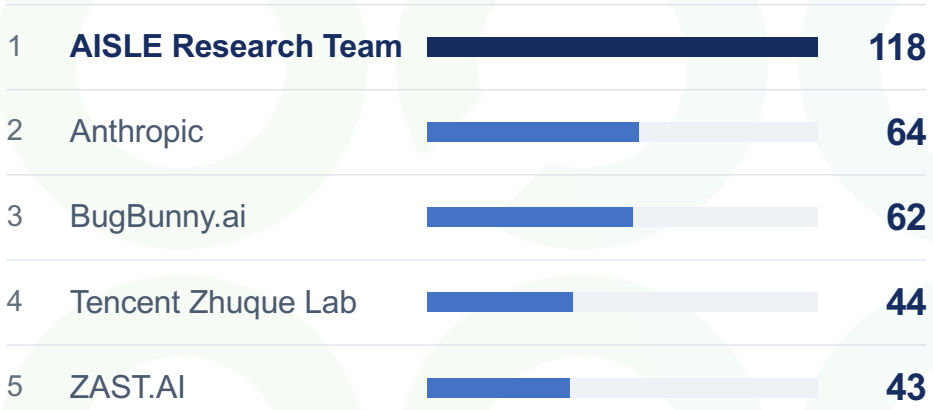
AISLE leads 5 of 8 Berkeley Vulnerability Initiative leaderboards.

Critical Count **41** • Anthropic	Avg Severity **8.49** • Xuanwu Atuin AI
Avg EPSS **12.14%** • Hacktron AI	CWE Breadth **118** • **AISLE Research Team**
MITRE Top-25 Reach **22** • **AISLE Research Team**	OWASP Reach **8** • **AISLE Research Team**
Volume **466** • **AISLE Research Team**	90d Finds **207** • **AISLE Research Team**

BVI Berkeley Vulnerability Initiative Source: vuln.cs.berkeley.edu

CWE BREADTH

Distinct CWE IDs touched · top 5 of 22 organizations with 10+ CVEs



ISACA Czech Republic Chapter

An AI model reports “no findings”. Can you trust it?

We used 10 models to find 95 real vulnerabilities. Each model did each test 10 times. Result: one scan is not sufficient. The most expensive model is not the best choice.

Presentation	https://arxiv.org/abs/2607.27030
Dataset	https://huggingface.co/datasets/aisleinc/HoF-Bench
Project	https://github.com/weareaisle/HoF-Bench

ISACA Czech Republic Chapter

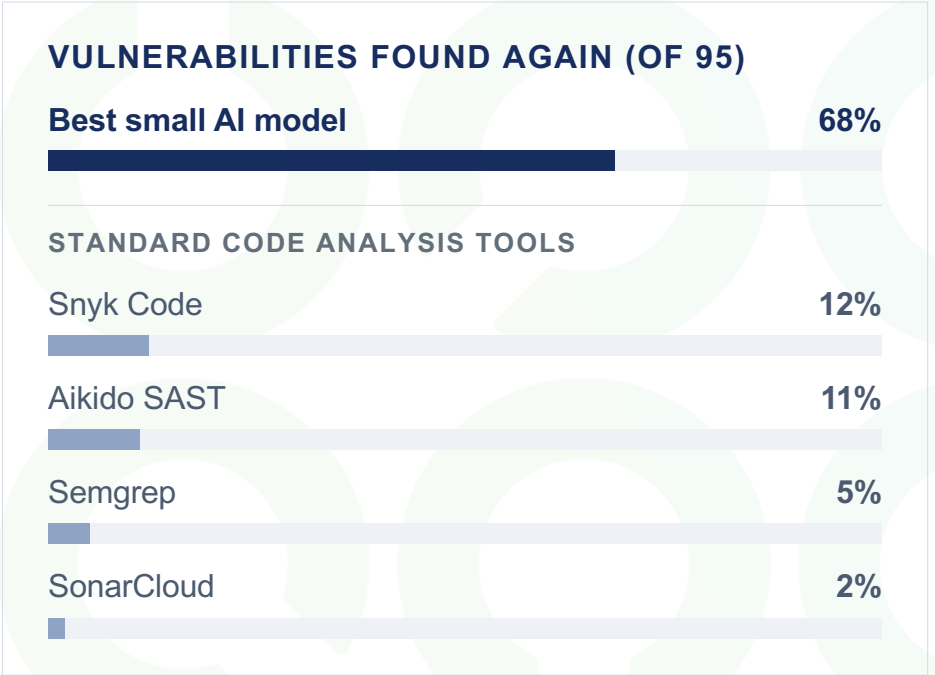
Small, low-cost AI models found most of the vulnerabilities.

68%

OF 95 VULNERABILITIES FOUND AGAIN BY THE BEST SMALL MODEL

65 of 95 · four runs · strict scoring

We did not use frontier AI. All ten models are small, fast and low-cost.



Standard tools: one indicative scan of the same code, with less strict scoring. Best model: GPT-5.6 luna. Source: AISLE HoF-Bench 2026

ISACA Czech Republic Chapter

Do the scan again. Also use a different model.

AI results change from run to run. Each model has blind spots. More runs and more models find more vulnerabilities.

ALL 95 VULNERABILITIES



47
One run
best model

65
Four runs
best model

68
One run each
four best models

84
Four runs each
all ten models

+12 to 20
more vulnerabilities for each model, with four runs instead of one

70 for $66
12 runs on three low-cost models find more than four runs of the best model (65 for $128)

72 of 95
found in one run by the full AISLE production system – the system that first found them

Vulnerabilities found, of 95, with generated context. “One run” is the average of all runs. Best model: GPT-5.6 luna. Source: AISLE HoF-Bench 2026

ISACA Czech Republic Chapter

Legacy SAST misses most real CVEs – AI-native analysis finds five times more.

Rule-based SAST only finds what someone wrote a rule for – and floods teams with noise.

MISSED BY SAST

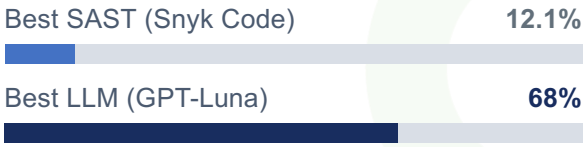
47–80%

of real-world CVEs missed by commercial SAST tools that score ~95% on synthetic benchmarks.

AI VS. SAST

5×

more CVEs found by AI analyzers than legacy scanners, on 95 public CVEs.



FALSE POSITIVES

<5%

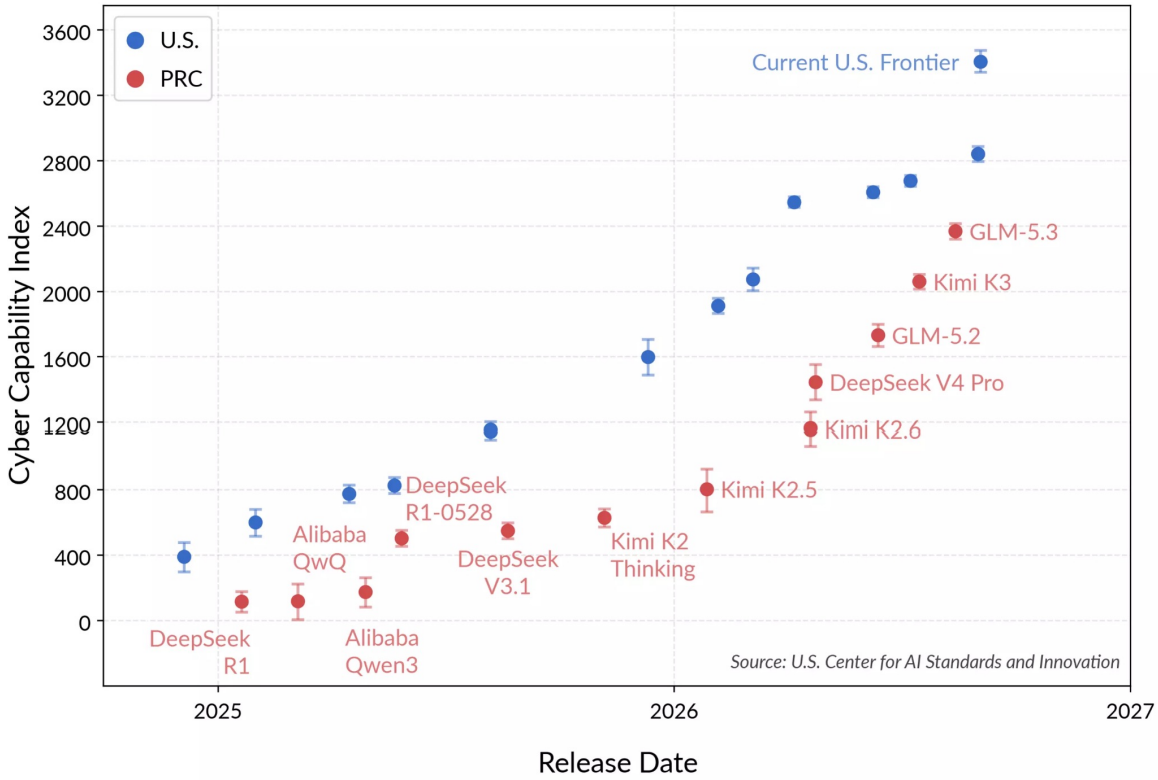
effective false-positive rate once every finding is verified – the point where engineers trust alerts again.

Sources: Lipp, Banescu & Pretschner, ISSTA 2022 · AISLE HoF-Bench 2026, arXiv:2607.27030

ISACA Czech Republic Chapter

4 months

The NIST estimate of how far GLM-5.3 is behind the U.S. frontier



Source: nist.gov/news-events/news/2026/09/caisis-assessment-zais-glm-53-cyber-capabilities

ISACA Czech Republic Chapter

Key Learnings

ISACA
Czech Republic Chapter

Detection is only the first link – AI must close the whole VulnOps chain.

Most breaches involving exploitation target flaws that already had a patch. More scanners don’t fix a broken workflow.

01

Analysis

Read code for intent and behavior. Rank real risks instead of listing thousands of findings.

02

Remediation

Generate fixes that follow your code patterns, so nothing breaks downstream.

03

Verification

Test every patch automatically. An unverified fix is a proposal, not a solution.

04

Improvement

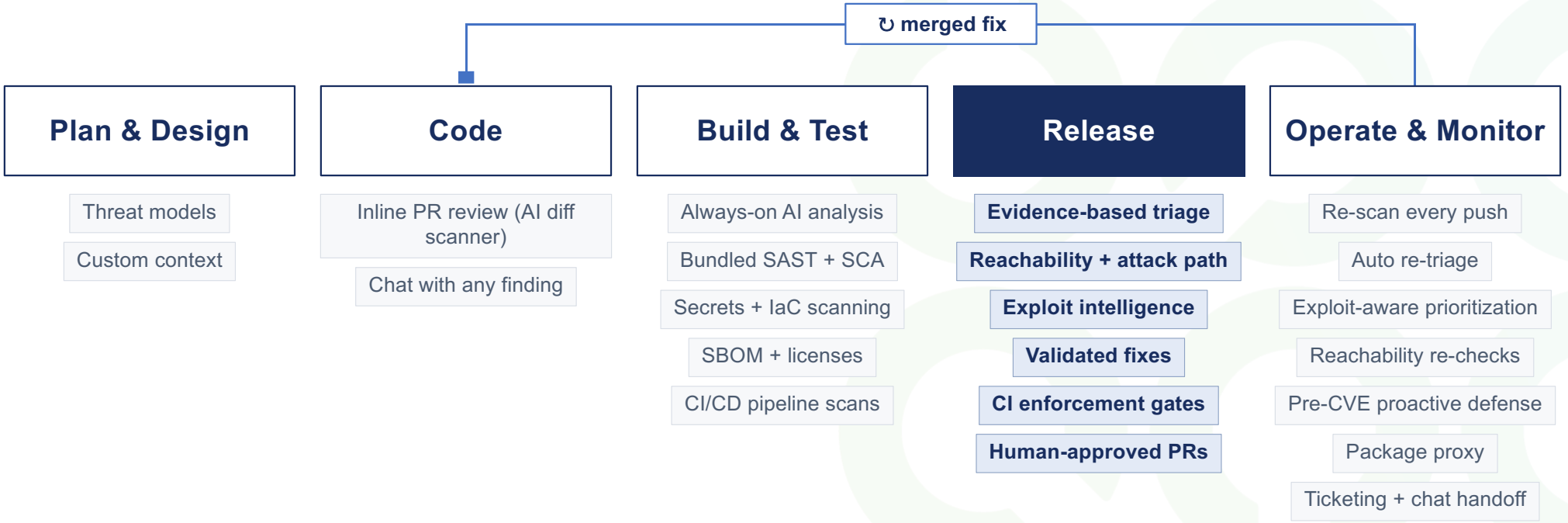
Deepen context as code ships, dependencies update and new vulnerability classes appear.

One unified flow – no gaps between analysis, remediation and verification, even on-prem or air-gapped.

Sources: Verizon 2025 DBIR · aisle.com/learn/sovereign-ai-cybersecurity

ISACA Czech Republic Chapter

Cover the whole SDLC



ISACA Czech Republic Chapter

Every company faces different cost and model constraints.

Small models alongside frontier models is the only mix that fits.

INPUTS · SET PER COMPANY

Price

Low — High

Availability

Restricted — Open

Speed

Batch — Real-time

MODEL MIX



- **Small models** — breadth
- **Frontier models** — depth

OUTPUT

Maximum coverage

within each company's limits

ISACA Czech Republic Chapter

Different security tasks need different levels of intelligence.

Match the model to the task.

INSTRUCTED SMALL MODELS

HIGH-INTELLIGENCE MODEL

UNCONSTRAINED AGENTS

FRONTIER MODELS

Scan

Identify which code is worth examining

Detect

Spot what's wrong in that code

Triage

Decide if it's a real bug or a false positive, and determine severity

Exploit

Turn the bug into a working attack

Patch

Write a fix that closes the bug without breaking the code

WHAT AISLE DOES

ISACA Czech Republic Chapter

How do you decide which model runs which task?

Evals decide.

Measure

Score every subtask on its own eval: scan, detect, triage, exploit.

Validate

Gate every change, to the system or the model, before it ships.

Enrich

Every failure in the field becomes a new test case. The dataset compounds, and becomes the moat.

ISACA Czech Republic Chapter

Deployment needs to fit the organization

Your data classification decides where the code can go and where the model can run.

FASTEST START

SaaS

Code leaves your perimeter

API models ZDR

YOUR CLOUD TENANT

Private Cloud

Inside your AWS, GCP or Azure

Models in your cloud tenant

YOUR DATA CENTER

On-Prem

Kubernetes, docker

Self-hosted open models

MAXIMUM ISOLATION

Air-Gapped

Fully isolated, zero egress

Local models only

← More convenience — More control →

ISACA Czech Republic Chapter

Results in Moneta

ISACA
Czech Republic Chapter

Summary of the assessment: No incident-worthy finding in Smart Banka; multiple areas can be hardened.

SUMMARY

- Our Android application is robust (for now)
- Exceptional support from AISLE during the assessment (important to reduce false positives)
- Thanks to rapid development of the AI, continuous service is needed
- “Standard” penetration tests for validation still required

ISACA Czech Republic Chapter

Risks

ISACA
Czech Republic Chapter

How do you solve this?

01 Geopolitics

02 Costs

03 Model poisoning

04 Sovereignty, ZDR

05 Models going rogue, cheating

ISACA Czech Republic Chapter

Thank you for having us!

ISACA
Czech Republic Chapter

Questions?

Pavel Bartoň
MONETA Money Bank, a.s.
pavel.barton@moneta.cz

Peter Koreň
AISLE
peter.koren@aisle.com

ISACA Czech Republic Chapter