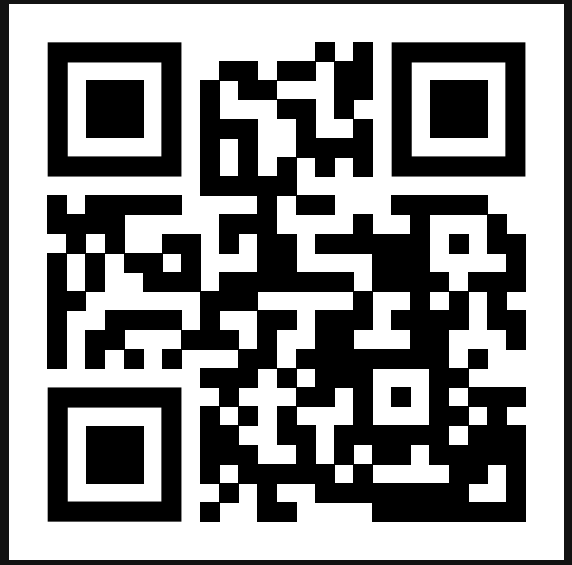
Bug Bounty Hunting with AI Agents

Can I automate bug bounty hunting using AI agents?

David Übelacker

👨‍💻 Who am I?

- **David ?belacker**
- Software Architect @ nag informatik ag in Basel
- 20+ years of experience in web and mobile application development



uebelacker.dev

nag
your IT-solution
our passion

What I dream of doing

What I’m actually doing

What is Bug Bounty Hunting?

Companies pay ethical hackers to find and report security vulnerabilities.

Popular Platforms

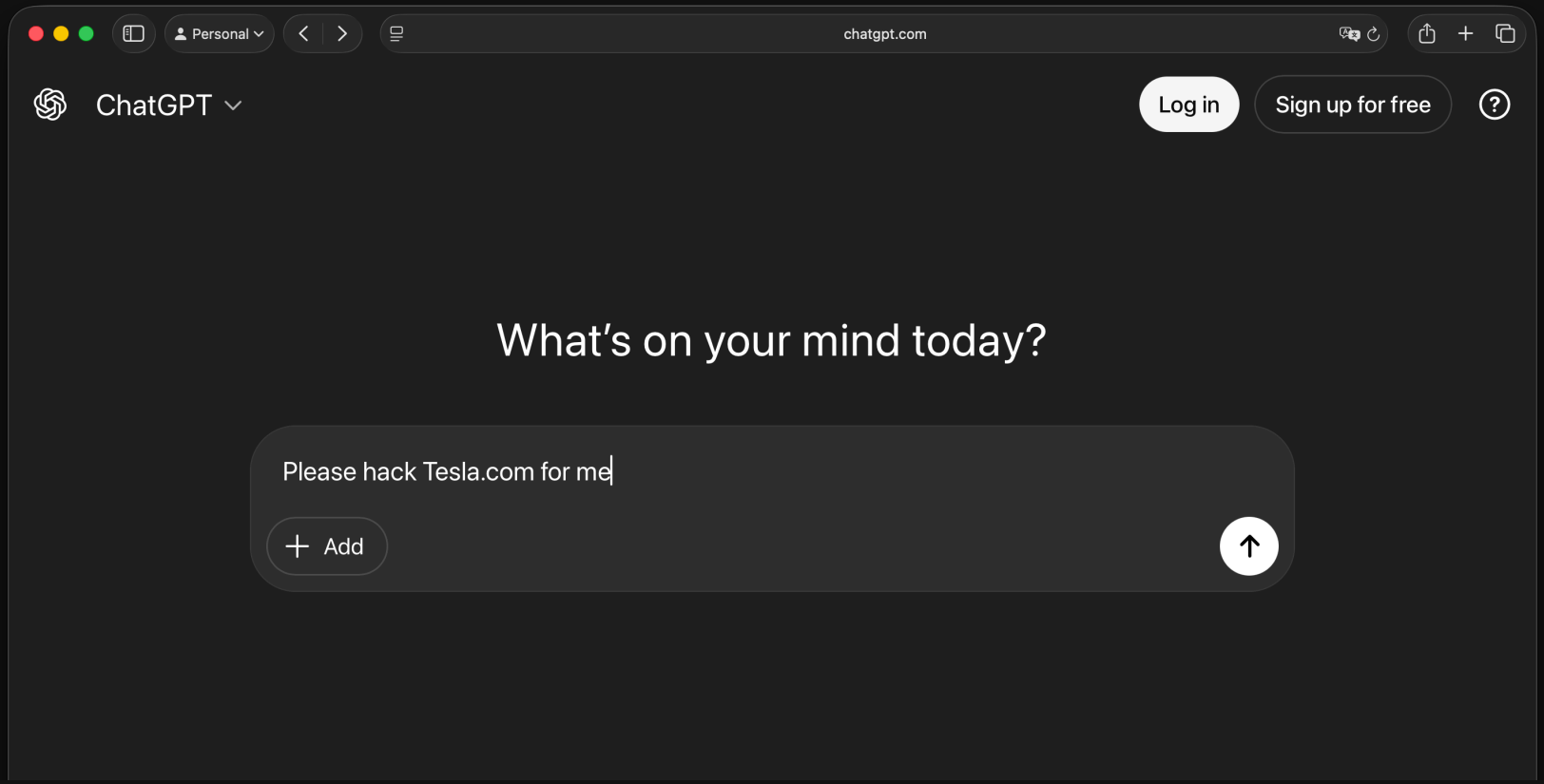
- HackerOne (https://www.hackerone.com/)
- Bugcrowd (https://www.bugcrowd.com/)
- Intigriti (https://www.intigriti.com/)
- 🇨🇭 Bug Bounty Switzerland (https://www.bugbounty.ch/)

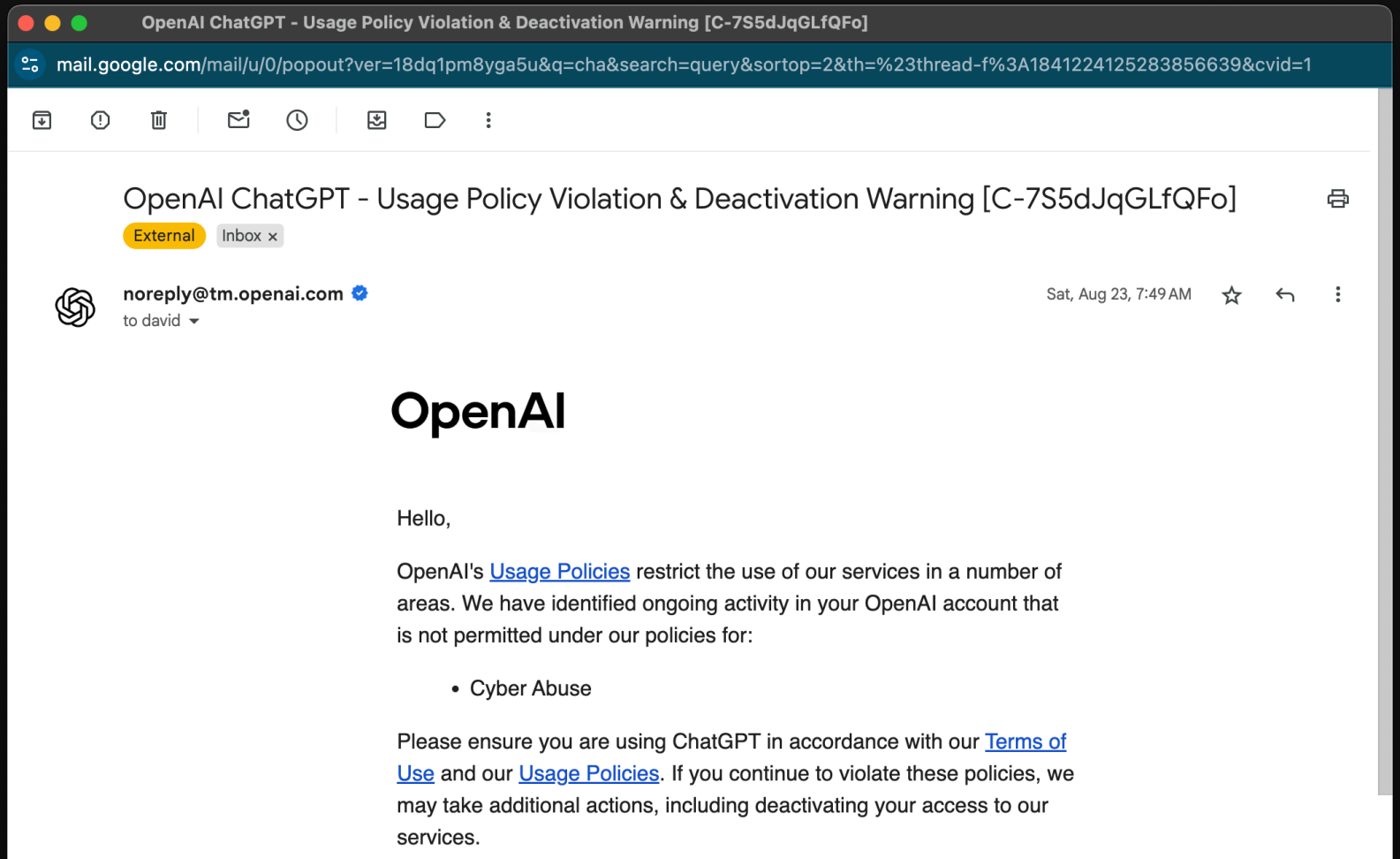
How to learn hacking

- Web security (OWASP - The Open Worldwide Application Security Project) https://owasp.org/
- Hack The Box https://www.hackthebox.com/
- Try Hack Me https://tryhackme.com/
- Ask ChatGPT

Can I automate bug bounty hunting using AI agents?

Attempt #1

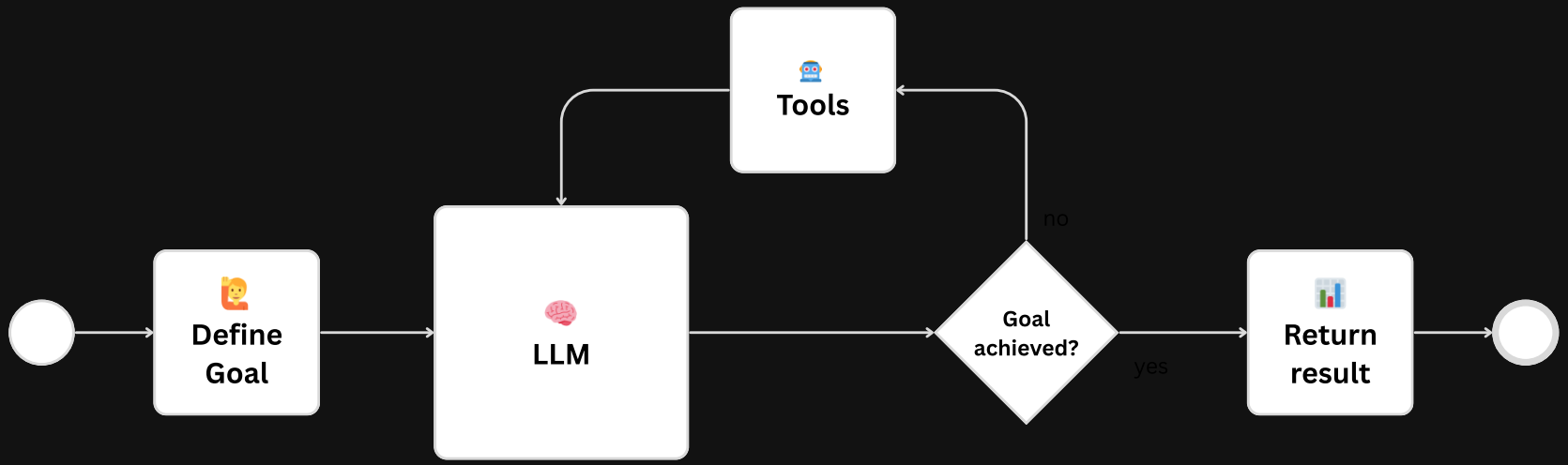

Personal
chatgpt.com
ChatGPT
Log in
Sign up for free
What's on your mind today?
Please hack Tesla.com for me
Add


OpenAI ChatGPT - Usage Policy Violation & Deactivation Warning [C-7S5dJqGLfQFo]
mail.google.com/mail/u/0/popout?ver=18dq1pm8yga5u&q=cha&search=query&sortop=2&th=%23thread-f%3A1841224125283856639&cvid=1
OpenAI ChatGPT - Usage Policy Violation & Deactivation Warning [C-7S5dJqGLfQFo]
External
Inbox ×
noreply@tm.openai.com
to david
Sat, Aug 23, 7:49 AM
OpenAI
Hello,
OpenAI's Usage Policies restrict the use of our services in a number of areas. We have identified ongoing activity in your OpenAI account that is not permitted under our policies for:
• Cyber Abuse
Please ensure you are using ChatGPT in accordance with our Terms of Use and our Usage Policies. If you continue to violate these policies, we may take additional actions, including deactivating your access to our services.

Attempt #2

What is an AI Agent?

An AI agent is a system that takes a goal, uses a large language model (LLM) and tools, and iterates until the goal is achieved.



🦜 LangChain & LangGraph

LangChain

A framework for building applications powered by LLMs.

- 🧠 Multiple LLM providers
- 📦 Document and vector stores
- 🛠️ External tools and APIs

LangGraph

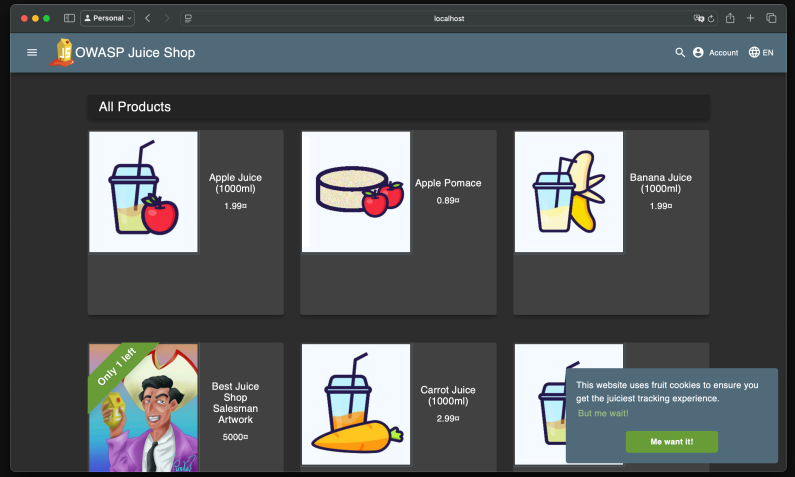
A framework for building complex, stateful AI agent workflows with advanced orchestration.

- 🗃️ **State management** - Persistent memory
- 🔵 **Nodes** - Workflow components
- ➡️ **Edges** - Conditional logic

Both are frameworks for Python, but there are equivalents for JavaScript / TypeScript (**LangChain.js**) and Java (**LangChain4j**).

OWASP Juice Shop

OWASP Juice Shop is a modern, insecure web app used for security training, with hacking challenges and as a 'guinea pig' for security tools.

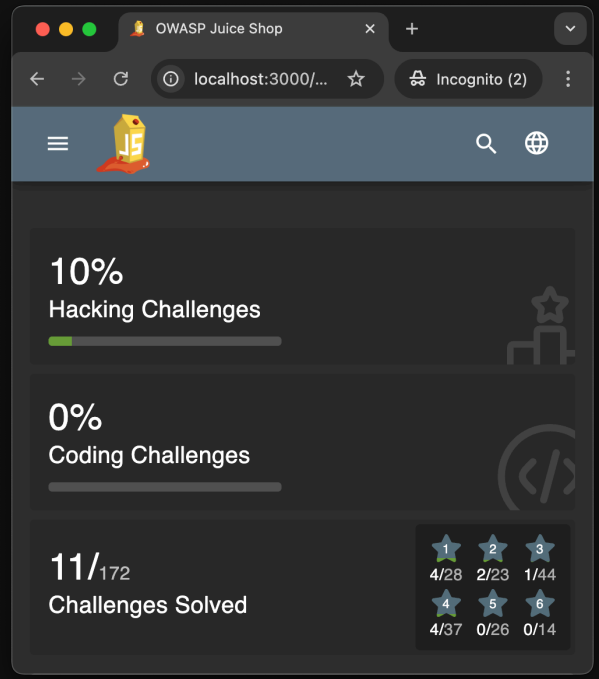


https://owasp.org/www-project-juice-shop/

Demo

Attempt #2 - Result

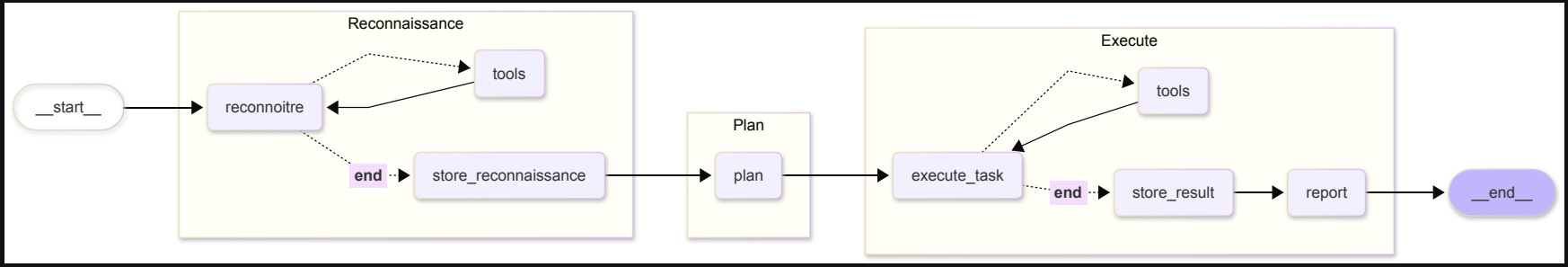
- Runtime: **2.5m**
- Tokens: **1.144.127**
- Costs: **3.47$**
- Hacking Challenges Solved: **11**



Attempt #3

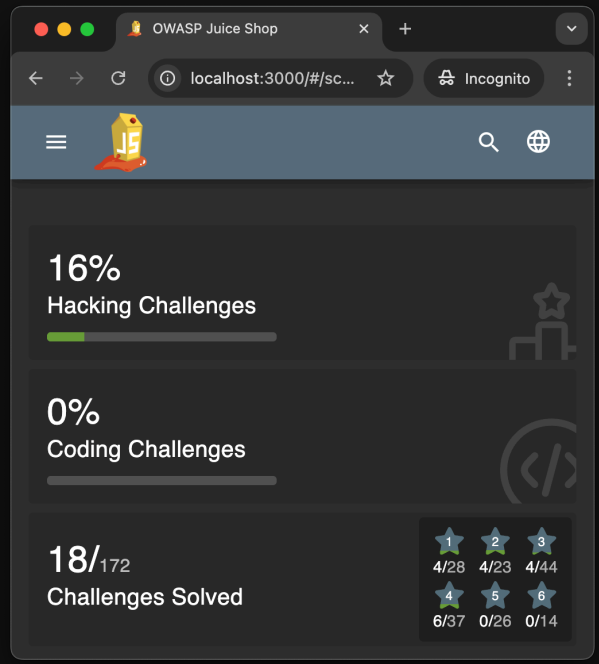
Three-phase workflow agent

- Phase 1: **Reconnaissance**
- Phase 2: **Plan**
- Phase 3: **Execute & Report**



Attempt #3 - Result

- Runtime: **73m**
- Tokens: **21’242’728**
- Costs: **65.59$**
- Hacking Challenges Solved: **18**



Cybersecurity AI (CAI)

Lightweight, open-source framework for AI-powered offensive & defensive automation. De facto AI Security framework, used by thousands of users & hundreds of organizations.

- 🤖 300+ AI Models (OpenAI, Anthropic, DeepSeek, Ollama, …)
- 🔧 Built-in security tools (reconnaissance, exploitation, privilege escalation)
- 🏆 Battle-tested (HackTheBox, bug bounties, real-world cases)
- 🎯 Agent-based modular architecture
- 🛡️ Guardrails: protection against prompt injection & dangerous commands
- 📚 Research foundation for democratizing Cybersecurity AI

https://github.com/aliasrobotics/cai
https://aliasrobotics.com/

Demo

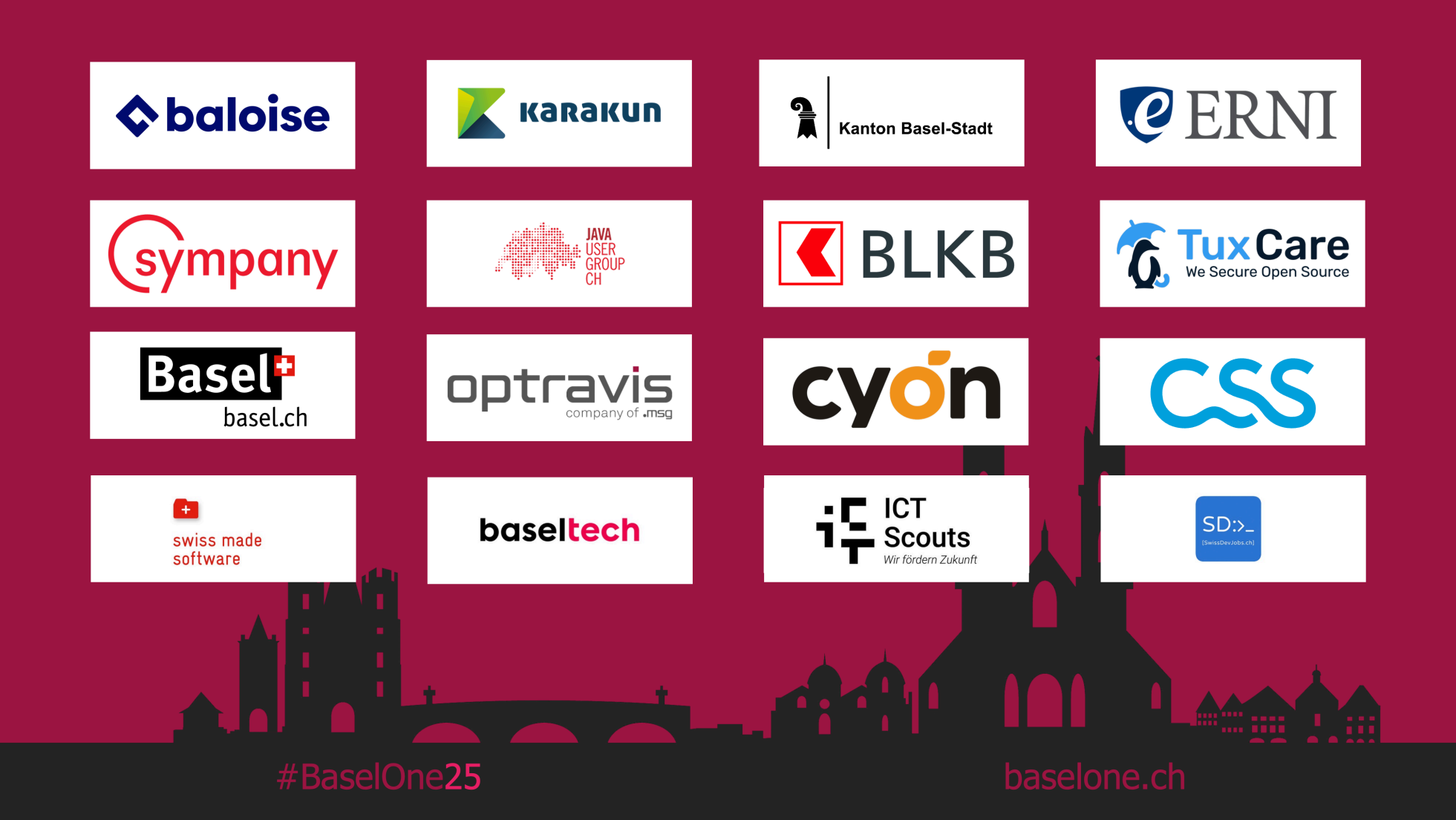
Key Takeaways

These key takeaways highlight both the opportunities and challenges of using AI in security.

- 💸 **Expensive** costs more than you earn in bounties
- 📏 **Context limits** — analysis of a lot of data runs fast into context limits
- 🔄 **Different models** — models behave differently, don’t expect the same results
- 🚨 **Easy** — unskilled hackers can launch easily AI-powered attacks
- ⚡ **Keep up** — security experts need to use AI in their daily work

Questions?

https://github.com/uebelack/bug-bounty-hunting-ai


baloise
KARAKUN
Kanton Basel-Stadt
ERNI
sympany
JAVA USER GROUP CH
BLKB
TuxCare
We Secure Open Source
Basel
basel.ch
optravis
company of msg
cyon
CSS
swiss made software
baseltech
ICT Scouts
Wir fördern Zukunft
SD:>_
[SwissDevJobs.ch]
#BaselOne25
baselone.ch