Equixly's AI Agents: Redefining Offensive API Security Testing

Alessio Dalla Piazza | Co-Founder & CTO, ***Equixly***

$ whoami - Alessio Dalla Piazza

Relatore Clusit

- Passion: inspired by the **RBT4** forum
- Cybersecurity Consulting (**15+ years**)
- Love breaking things. CVEs (Apples Safari, VMWare, IBM WebSphere, Docker…)
- Co-Founder of **Equixly | AI-Powered API Security Testing Platform**

- Fixed the permissions on `%PROGRAMDATA%\Docker` to avoid a potential Windows containers compromise. See CVE-2021-37841. Thanks to Alessio Dalla Piazza for discovering the issue.

Release … ↑ Acknowledgement

2023

Nov 30, 2023 Alessio Dalla Piazza with Equixly

CVSS Vector: (CVSS:3.0/AV:N/AC:H/PR:L/UI:N/S:U/C:L/I:H/A:L)

Affected Products and Versions

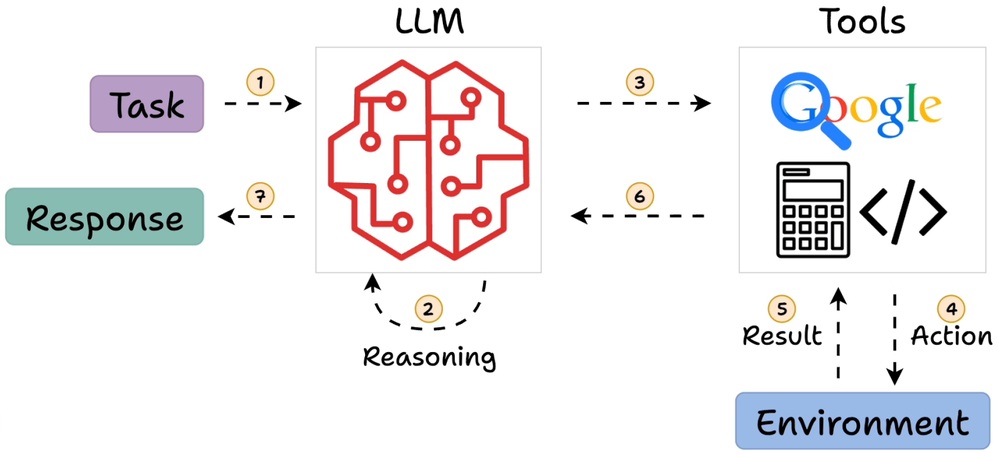
Affected Product(s)	Version(s)
WebSphere Application Server ND	9.0
WebSphere Application Server ND	8.5

Acknowledgement

This vulnerability was reported to IBM by Alessio Dalla Piazza.

$ What is an “Agent”? A “loop” that ask a model a question…

- LLMs provide **static responses** (pure text or answers)
- They have **limited capability**, lacking environmental interaction
- **Agents** enhance LLMs by enabling interaction with tools and environmental feedback
- Agents allow LLMs to **iteratively** solve complex problems through dynamic decision-making *aka try-harder*

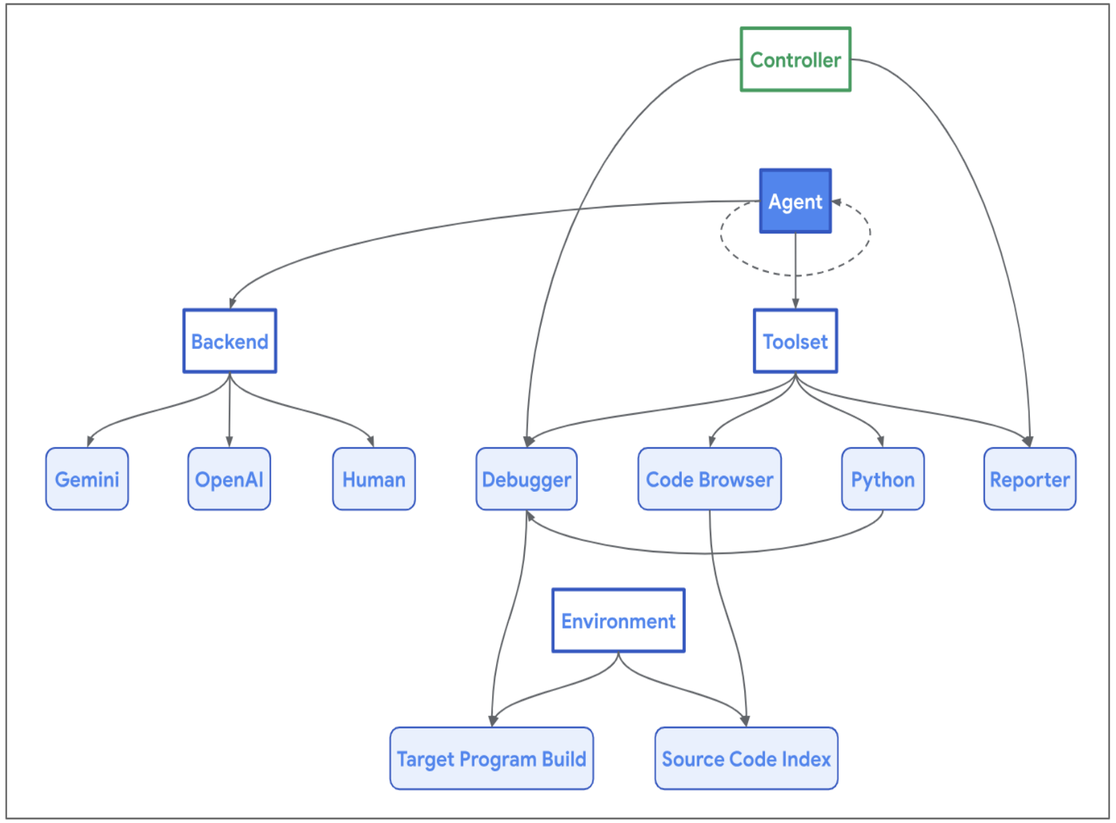


$ Tool Use – As humans use a calculator, models need too

- LLMs excel at language understanding and generation but struggle with computations
- Like humans who rely on calculators, LLMs benefit from external tools to perform calculations too

$ Google Project Naptime – why using an Agent helps

- Meta introduced **CyberSecEval2** to measure LLMs security capabilities
- Initial eval appeared to show very low performance
- Google used an agent with a "question-in-loop" approach, improving success rate from 0.05 to 1 in **buffer overflow**

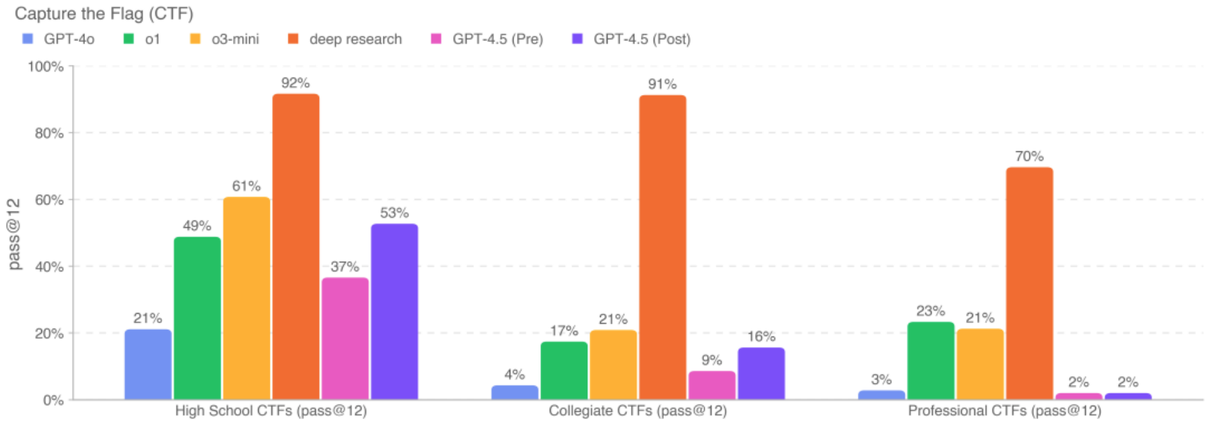


Source: https://googleprojectzero.blogspot.com/2024/06/project-naptime.html

$ LLMs for Offensive Security

GPT-4.5 Vulnerability Exploitation Score: Low

- 53% success in high-school level CTF challenges
- 16% in collegiate
- 2% in professional CTFs
- Does not meet medium-risk threshold for real-world vulnerability exploitation



Source: https://cdn.openai.com/gpt-4-5-system-card-2272025.pdf

$ What does this mean for offensive cybersecurity?

- **AI Agents Rise**: LLMs now act autonomously in environments
- **Commercial AI Advancements**: Driven by competition and investment (e.g., Gemini, Claude, GPT)
- **Open-Source Progress**: Distillation improves models like Deepseek, QWEN, LLAMA…
- **Cybersecurity Concerns**: Increased focus from AISI, NIST, AI ACT on AI risks

Example: https://www.forbes.com/sites/johnwerner/2025/01/30/did-deepseek-copy-off-of-openai-and-what-is-distillation/

$ Are Existing Tools Enough for humans to Effectively Test APIs?

Short answer: **NO**

- **Limited Coverage:** Humans solved less than half the API challenges – none of them using tools
- **High Error Rates:** Automated tools used by humans caused **80,000+** failed requests
- **"Spray-and-Pray" Methodology:** Traditional tools lack context-awareness, leading to ineffective tests
- **Minimal Manual Exploration:** Less than 1% of testing was genuinely manual
- **AI Advantage: Equixly AI identified 230 vulnerabilities in 1 hour**, outperforming manual testing

$ Are Existing Tools Enough for AI Agents to Effectively Test APIs?

Short answer: **NO**

- **AI Agents vs. Human-Centric Tools**: Tools like ZAP, Burp are not built for automated AI workflows
- **Tool Limitations**: Limited APIs and automation hinders AI agent (not born for programmatic usage)
- **LLM Capabilities Unused Potential**: Existing tools restrict intelligent decision-making
- **Need for Dynamic Tools**: AI agents require context-aware, dynamic API testing tools
- **Must Have AI-Native Tools**: Only AI-native tools unlock full potential for LLM-driven testing

$ Equixly: AI-Powered Agent for API Security Vulnerability Detection

- **Continuous Scanning:** Regular AI-powered API monitoring
- **OWASP Top 10 Testing**
- **Attack Surface Mapping:** Inventories API ops and data flows
- **Compliance Reporting:** Highlights security risks and data exposure
- **Advanced AI Techniques:** Detects zero-day vulnerabilities and shadow APIs

$ Equixly: The Power of Smart Fuzzing

Traditional Fuzzing

- **Randomized Inputs:** Employs arbitrary data inputs without system-specific tailoring
- **Static Testing:** Lacks adaptability, potentially missing complex vulnerabilities
- **Generic Approach:** Misses specific attack scenarios by generating simple strings e.g., SSN

Smart Fuzzing

Used by Equixly

- **Intelligent Input Generation:** Crafts inputs based on target system insights
- **Iterative Learning:** An AI agent refines testing strategies by learning from previous responses
- **Customized Language Models:** Utilizes tailored LLMs to simulate real-world attack vectors effectively

$ Equixly: Context Awareness

- **Context-Aware Testing:** Equixly understands the purpose of each endpoint as it creates security tests
- **False Positive Detection:** Equixly understands context (e.g., knowing "blog/1" is ok to be publicly accessible whereas "orders/1” is not), reducing false alerts
- **Accurate Detection:** Recognizes specific paths and permissions, identifying hidden access points
- **Tailored Approach:** Focuses on the unique parts of an app to match real-world attack scenarios

$ Equixly: Simplified AI Integration & Data Marshalling

- **Custom AI Functions:** Easy integration with internal tools using custom/native AI functions
- **Data Marshalling:** Ensures clear inputs and outputs for the LLM
- **Seamless Communication:** Facilitates smooth interaction with other algorithms
- **Improved Accuracy:** Organized data flow leads to more accurate results

$ Equixly: Simplified AI Integration & Data Marshalling

#		Path
1	POST	/microservicebola/register
2	POST	/microservicebola/register
3	POST	/microservicebola/login
4	POST	/microservicebola/documents
5	POST	/microservicebola/documents
6	POST	/microservicebola/documents
7	POST	/microservicebola/documents
8	POST	/microservicebola/documents
9	POST	/microservicebola/documents
10	POST	/microservice...rders/create
11	GET	/microservice...s/index.html
12	GET	/microservice...munity/forum
13	GET	/microservice...56be6de93a2d
14	GET	/microservice...r/2926135474
15	GET	/microservice.../api/order/2
16	POST	/microservice...l/createteam

Request

```
1  POST /microservicebola/register HTTP/1.1
2  Host: local.equixly.com:5656
3  Accept: text/plain, application/json
4  Content-Type: application/json
5  User-Agent: Equixly/1.0 (API Security; ML-powered Testing)
6  X-Equixly-Auth-Profile: <nil>
7  X-Equixly-Sequence-Namespace: 1d681623-0fe8-4ff3-b665-5fe569938b62
8
9  {
10 "name": "Ann",
11 "password": "oiFD+jWU&9"
12 }
```

Response

```
1  HTTP/1.1 400 Bad Request
2  Content-Length: 115
3  Content-Type: application/json; charset=utf-8
4  Date: Wed, 05 Mar 2025 11:05:13 GMT
5
6  {
7  "error": "Key: 'BOLARegisterAccountRequest.Email' Error:Field validation for 'Email' failed on the 'required' tag"
8  }
```

$ Equixly: Simplified AI Integration & Data Marshalling

#	Path	
1	POST	/microservicebola/register
2	POST	/microservicebola/register
3	POST	/microservicebola/login
4	POST	/microservicebola/documents
5	POST	/microservicebola/documents
6	POST	/microservicebola/documents
7	POST	/microservicebola/documents
8	POST	/microservicebola/documents
9	POST	/microservicebola/documents
10	POST	/microservice...rders/create
11	GET	/microservice...s/index.html
12	GET	/microservice...munity/forum
13	GET	/microservice...56be6de93a2d
14	GET	/microservice...r/2926135474
15	GET	/microservice.../api/order/2
16	POST	/microservice...l/createteam

Request

```
1  POST /microservicebola/register HTTP/1.1
2  Host: local.equixly.com:5656
3  Accept: text/plain, application/json
4  Content-Type: application/json
5  User-Agent: Equixly/1.0 (API Security; ML-powered Testing)
6  X-Equixly-Auth-Profile: <nil>
7  X-Equixly-Sequence-Namespace: 1d681623-0fe8-4ff3-b665-5fe569938b62
8
9  {
10 "email": "patrickgibson@equixly.com",
11 "name": "Eric",
12 "password": "J!S6Z2jdoD"
13 }
```

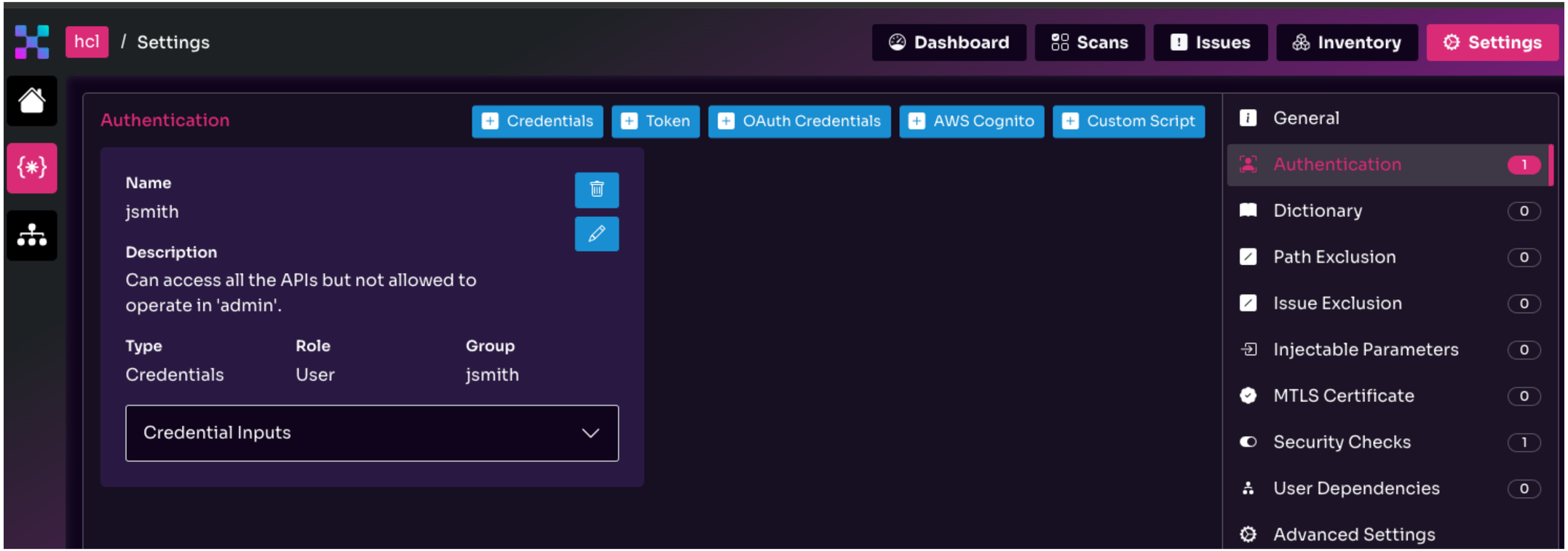
Response

```
1  HTTP/1.1 200 OK
2  Content-Length: 31
3  Content-Type: application/json; charset=utf-8
4  Date: Wed, 05 Mar 2025 11:05:14 GMT
5
6  {
7  "Status": "Success",
8  "UserID": 8
9  }
```

$ Equixly: From Functional Analysis to RBAC

- **Human Language to RBAC Matrix:** Translates functional analysis in plain language into a clear RBAC matrix
- **Identifying Security Issues:** Helps detect RBAC security flaws by mapping roles and permissions
- **The Future of Security:** Automated, intelligent analysis for better security management

$ Equixly: From Functional Analysis to RBAC



$ Equixly: From Functional Analysis to RBAC

hcl / Scans / Scan Detail • Running

Dashboard | Scans | Issues | Inventory | Settings

General | Issues 2 | Authorization Matrix | HTTP History | Endpoints | Dependencies Graph | Settings

Path

Endpoint	Severity	Created	Name	Category	OWASP
POST /admin/changePassword	High	Sun Jan 26, 2025	Authorization Schema Misconfiguration	Stored Cryptography	API2:2023
POST /admin/addUser	High	Sun Jan 26, 2025	Authorization Schema Misconfiguration	Stored Cryptography	API2:2023

2 items 1 12 / page

$ Attackers Only Need to Win Once; Defenders Must Win Continuously

- Offensive security is advancing rapidly with increasingly powerful AI agents
- To effectively combat attackers, we must also build and understand attacks ourselves
- Companies should have access to top-tier attack tools to identify vulnerabilities before software is released
- Recent policy asks, which aim to restrict models from helping with exploits, are misguided
- We need to develop robust AI-based tools for offensive security and empower developers to fix vulnerabilities during the development process

Q&A

Contatti:

alessio.dallapiazza@equixly.com

Vieni a trovarci al nostro stand!