HADRIAN

THE OPEN-SOURCE AI ATTACK TOOLKIT

87 tools. 18 months. Threat actors already exploiting.

The open-source AI pentesting ecosystem barely existed before April 2023. Today Hadrian has catalogued 87 tools across seven categories — autonomous agents, guardrail-free offensive LLMs, reverse engineering frameworks, and recon platforms — the majority launched in the 18 months since GPT-4's release. One of them, CyberStrikeAI, was weaponised in attacks across 55 countries within weeks of publication.

Key statistics

87	<5	140K	<8 WKS
TOOLS CATALOGUE MARCH 2026	EXISTED BEFORE GPT-4(APR 2023)	DOWNLOADS GUARDRAIL-FREE LLM	PUBLISH → WEAPONISED CYBERSTRIKEAI · JAN 2026

The inflection point — before and after GPT-4

BEFORE GPT-4 · PRE APRIL 2023

~8

OPEN-SOURCE AI PENTESTING TOOLS EXISTED

TODAY · APRIL 2026

79

TOOLS CATALOGUED IN SINCE GPT-4 WAS RELEASED

Category definition

Tools are grouped into seven categories reflecting their primary capability and target. A tool may have secondary applications; the primary use determines placement.

- **AUTONOMOUS**

  Full-pipeline agents that conduct penetration tests end-to-end with minimal human involvement — recon, scanning, exploitation and reporting all orchestrated by AI. These represent the highest-capability tier: tools that replace, not assist, the human tester.

- **VULN**

  Tools that use AI to discover, classify, or exploit software vulnerabilities — including fuzzing, static analysis, CVE matching, and AI-generated exploit code. Distinct from Autonomous tools in that they target a specific phase rather than the full kill chain.

- **REV.ENG.**

  AI-assisted reverse engineering tools that analyse compiled binaries, decompiled code, or malware. Capabilities include function renaming, vulnerability detection in decompiled output, and automated code explanation — dramatically accelerating malware analysis and binary exploitation.

RECON

AI-powered intelligence-gathering and attack-planning tools — subdomain enumeration, CVE correlation, threat modelling, Burp Suite integrations, and conversational interfaces for building target profiles. Lowers the entry barrier for attack planning.

LLM RT

LLM Red-Teaming tools designed to probe, stress-test, and attack AI systems themselves — jailbreaking, prompt injection, adversarial inputs, and safety evaluation frameworks. As AI is embedded in enterprise systems, this category is increasingly relevant to enterprise attack surface.

CTF

AI agents built or optimised for Capture the Flag competitions. CTF tools serve as proving grounds for autonomous exploit capability: several have solved challenges requiring multi-step reasoning, binary exploitation, and web vulnerabilities without human guidance.

CYBER LLM

Purpose-built or fine-tuned language models trained on offensive security data, malware corpora, and hacking knowledge — with guardrails deliberately removed or absent. These answer attack methodology questions directly, generate working exploit code, and run locally with no logging.

Complete tool catalog - 87 tools

Published = first public GitHub release | Stars / downloads as of April 2026

TOOLS	PUBLISHED	STARS / DL	CATEGORY
Strix	SEP 2025	~20k	AUTONOMOUS
PentestGPT	MAR 2023	~11.9k	AUTONOMOUS
Shannon	FEB 2026	~10.6k	AUTONOMOUS
PentAGI	JAN 2026	~8.9k	AUTONOMOUS
CAI	JUN 2025	~7.2k	AUTONOMOUS
HexStrike AI	APR 2025	~7.2k	AUTONOMOUS
PentestAgent	MAR 2025	~1.7k	AUTONOMOUS
⚠ CyberStrikeAI	NOV 2025	~1.3k	AUTONOMOUS
RedAmon	Q2 2024	~1.4k	AUTONOMOUS
hackingBuddyGPT	JAN 2023	~965	AUTONOMOUS
NeuroSploit	Q3 2024	~870	AUTONOMOUS
Nebula	AUG 2024	~867	AUTONOMOUS
RamiGPT	JAN 2025	~838	AUTONOMOUS
Reaper	JUL 2024	~831	AUTONOMOUS
Rogue	FEB 2025	~448	AUTONOMOUS

TOOLS	PUBLISHED	STARS / DL	CATEGORY
Pentest Copilot	SEP 2024	~233	AUTONOMOUS
HackSynth	DEC 2024	~278	AUTONOMOUS
Floki	APR 2024	~200+	AUTONOMOUS
VulnBot	JAN 2025	~100+	AUTONOMOUS
AutoPentester	Q3 2024	~100+	AUTONOMOUS
AI-OPS	AUG 2024	~111+	AUTONOMOUS
Cochise	Q3 2024	~100+	AUTONOMOUS
SecGPT	Q1 2023	~200+	AUTONOMOUS
Decepticon	MAR 2025	~50+	AUTONOMOUS
TARS	APR 2025	~30+	AUTONOMOUS
AutoPentest	MAY 2025	~50+	AUTONOMOUS
BlacksmithAI	MAR 2026	~103	AUTONOMOUS
WRN-PentestGPT	Q4 2024	~39+	AUTONOMOUS
Cyber-AutoAgent	NOV 2025	~504	AUTONOMOUS
ARTEMIS	DEC 2025	~489	AUTONOMOUS
MetasploitMCP	2025	~521	AUTONOMOUS
EVA	FEB 2026	~385	AUTONOMOUS
MAPTA	AUG 2025	~98	AUTONOMOUS
Deadend CLI	Q1 2026	~200+	AUTONOMOUS
CortexAI	2025	~12	AUTONOMOUS
Vulnhuntr	OCT 2024	~2.5k	VULN
PassGAN	AUG 2017	~4k	VULN
oss-fuzz-gen	OCT 2023	~1.3k	VULN
FuzzyAI	DEC 2024	~1.2k	VULN
GPT_Vuln-analyzer	MAY 2023	~1k+	VULN
nuclei-ai-ext	Q3 2024	~500+	VULN
AutorizePro	Q3 2024	~300+	VULN
EscalateGPT	JUN 2023	~118	VULN
vulchatgpt	JUL 2023	~80+	VULN
Ret2GPT	AUG 2023	~60+	VULN

TOOLS	PUBLISHED	STARS / DL	CATEGORY
LLM4Decompile	MAR 2024	~6.1k+	REV.ENG.
Gepetto	DEC 2022	~3.4k	REV.ENG.
ReverserAI	MAR 2024	~962	REV.ENG.
gpt-wpre	MAR 2023	~600+	REV.ENG.
ReVa	AUG 2023	~516	REV.ENG.
IATelligence	JUN 2023	~400+	REV.ENG.
G-3PO/Ghidra	MAY 2023	~400+	REV.ENG.
GhidrAssist	2024	~568	REV.ENG.
GhidrAssistMCP	FEB 2026	~523	REV.ENG.
Rikugan	MAR 2026	~274	REV.ENG.
Binary Ninja MCP	2025	~263	REV.ENG.
OGhidra	2025	~125	REV.ENG.
GhidraGPT	SEP 2023	~100+	REV.ENG.
BinAssist	2024	~50	REV.ENG.
BinAssistMCP	2025	~22	REV.ENG.
Burp AI Agent	2025	~775	RECON
MCP-Kali-Server	2025	~567	RECON
BurpGPT	FEB 2023	~2.3k	RECON
hackGPT	DEC 2022	~1.2k	RECON
STRIDE-GPT	JUN 2023	~983+	RECON
ReconAIzer	JUN 2023	~884	RECON
HackerGPT-2.0	SEP 2023	~800+	RECON
LLMFuzzer	JUL 2023	~341	RECON
Subwiz	NOV 2024	~348	RECON
SubGPT	SEP 2023	~200+	RECON
ChatCVE	OCT 2023	~100+	RECON
promptfoo	APR 2023	~10.8k	LLM RT
NVIDIA garak	JUN 2023	~4.7k	LLM RT
PurpleLlama	DEC 2023	~3.9k	LLM RT
PyRIT	FEB 2024	~3.4k	LLM RT

TOOLS	PUBLISHED	STARS / DL	CATEGORY
AI-Infra-Guard	OCT 2024	~1k+	LLM RT
agentic_security	AUG 2024	~500+	LLM RT
project_mantis	AUG 2024	~300+	LLM RT
InterCode-CTF	JUN 2023	~150+	CTF
D-CIPHER/NYU CTF	AUG 2024	~200+	CTF
Cyber-Zero	JAN 2025	~100+	CTF
llm-ctf-agent	Q3 2024	~20+	CTF
WhiteRabbitNeo	NOV 2023	140k DL	CYBER LLM
Foundation-Sec-8B	MAR 2025	~100k DL	CYBER LLM
AutoAudit LLM	SEP 2023	~400+	CYBER LLM
BaronLLM	Q4 2024	~141+	CYBER LLM
Lily-Cyber-7B	Q4 2024	~133+	CYBER LLM

The weaponization timeline - CyberStrikeAI

KEY FINDINGS: < 8 WEEKS

CYBERSTRIKEAI PUBLISHED NOVEMBER 2025 · THREAT ACTOR INFRASTRUCTURE OBSERVED JANUARY 2026

DATE	EVENT	DETAIL	KEY METRIC
NOVEMBER 2025	CyberStrikeAI published	100+ offensive tools, AI orchestration, YAML attack recipes, MCP support. Free on GitHub.	-
JANUARY 2026	First threat actor deployments	21 unique IPs running active CyberStrikeAI infrastructure identified within weeks.	21 IPs
JAN - FEB 2026	FortiGate campaign confirmed	CyberStrikeAI confirmed in attacks against Fortinet appliances across 55 countries.	55 Countries
2003 - 2020	Metasploit, Cobalt Strike	Each took years from first release to widespread attacker adoption.	Years → Weeks

What security leaders should do differently

#	RECOMMENDATION	RATIONALE	URGENCY
1	Assume the skill floor has collapsed	Shannon achieves 96% autonomous exploit success. Threat models built around attacker expertise requirements are outdated.	IMMEDIATE
2	Treat Publication as a Weaponization Trigger	Monitor the AI offensive tool ecosystem as actively as CVEs. CyberStrikeAI proved the window is weeks.	IMMEDIATE
3	Reassess External Attack Surface Exposure	Shannon, Strix, and HexStrike AI are optimised for web apps, APIs, and edge infrastructure. Any exposed service is at risk.	NEAR-TERM
4	Monitor the MCP Shift	Any AI agent can now invoke Kali tools directly via MCP. The threat surface includes every AI-connected environment.	NEAR-TERM
5	Audit Guardrail-Free LLMs in Your Environment	WhiteRabbitNeo (140k+ downloads) runs locally with no logging. If it is inside your environment, your posture is affected.	MEDIUM-TERM
6	Match Attacker Velocity with Continuous Testing	Weaponization: years to weeks. Point-in-time pentests are structurally insufficient against this cadence.	STRATEGIC