selfhack*

TECHNICAL WHITEPAPER

SelfHack AI: A New Era in Autonomous Penetration Testing

SelfHack AI is an autonomous penetration testing platform that emulates real-world attacker behavior using advanced AI models. It continuously scans, exploits, and prioritizes security vulnerabilities without manual input, helping organizations strengthen their defenses proactively, efficiently, and at scale.

ADDRESS

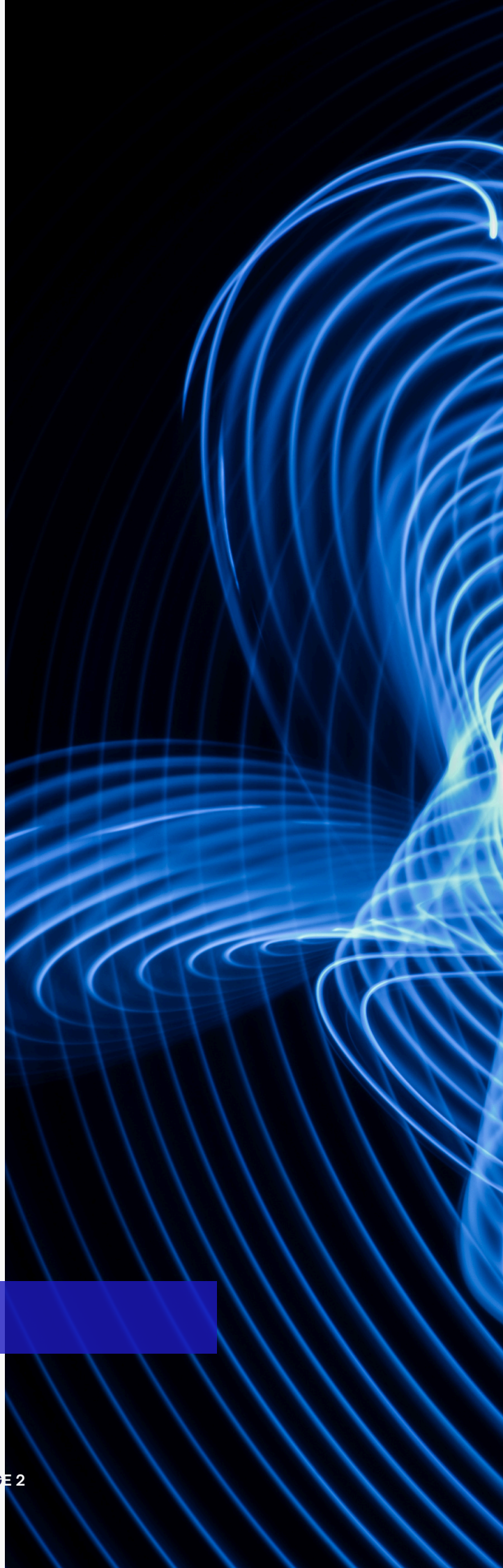
Maria 01
Helsinki, Finland

CONTACT US

info@selfhack.fi
www.selfhack.fi

Content

Click on the topics below to access the relevant page.



.01
Executive Summary

> **"What used to take weeks of expert analysis now takes minutes without sacrificing depth or accuracy."**

Rethinking Penetration Testing in the Age of AI

SelfHack AI is an autonomous cybersecurity platform that reimagines how organizations conduct penetration testing. Traditionally, penetration testing has relied on manual work by highly skilled security professionals a time-consuming, expensive, and infrequent process that leaves companies exposed between test cycles. SelfHack AI replaces this outdated model with a fully automated, AI-driven system capable of simulating real-world attacks, discovering complex security flaws, and delivering strategic and technical insights, all in a fraction of the time and cost.

At its core, SelfHack AI leverages intelligent agents that think and act like human attackers. These agents don't just scan for known vulnerabilities; they understand your environment's business logic, adapt to its architecture, and execute multi-stage attack chains to uncover weaknesses that are often invisible to traditional tools. Using a combination of symbolic reasoning, natural language processing, and continuous learning, our agents mimic adversarial behavior to expose not only CVEs, but also logic flaws, privilege escalations, and chained attack paths across web, mobile, API, and internal networks.

This shift from passive detection to active exploitation transforms cybersecurity from reactive to proactive, and from periodic to continuous. SelfHack AI enables organizations to test more often, respond more quickly, and gain a real understanding of how threats could unfold in their specific environment. By integrating seamlessly with existing DevSecOps, ticketing, and incident response systems, the platform enhances both offensive and defensive workflows, creating a unified approach to cyber resilience.

From a business perspective, SelfHack AI reduces operational costs, shortens time-to-remediation, and ensures consistent, audit-ready reporting aligned with compliance frameworks like ISO 27001, SOC 2, and GDPR. For technical teams, it eliminates noise by validating vulnerabilities through real-world exploitation and providing actionable recommendations tied to risk and business impact.

In a threat landscape where speed, complexity, and creativity define modern attacks, SelfHack AI introduces a new standard, scalable, intelligent, and autonomous penetration testing built for today's enterprise environments.

Introduction & Problem Statement

Cyber threats evolve by the hour. Traditional security testing doesn't.

Cybersecurity is at a breaking point. While attack surfaces expand rapidly through APIs, microservices, mobile apps, and remote infrastructure, the average organization continues to rely on periodic manual penetration testing. These are typically conducted once or twice a year, a frequency that no longer reflects the pace of digital change.

This delay leaves security teams blind to emerging vulnerabilities, business logic flaws, and chained attack paths that evolve daily. Reactive approaches are no longer enough. What's needed is a dynamic, continuous, and intelligent way to test security, as fast and adaptive as the attackers themselves.

SelfHack AI is built to close this gap. It transforms penetration testing from a manual, one-time task into a proactive, ongoing process. Using AI-driven agents that simulate real-world adversaries, it goes beyond checklists to think, learn, and act autonomously, just like a human red team would, but at machine scale and speed.

88%

of organizations say their security posture is reactive rather than proactive, with limited ability to anticipate how attackers might exploit complex, chained vulnerabilities within their unique environments.[1]

With SelfHack AI, testing becomes more than vulnerability detection, it becomes a strategic layer of defense, built on automation, contextual awareness, and adversarial thinking.

Key problems faced by modern organizations:

- Infrequent testing leads to blind spots and exposure to new threats
- Weak testing coverage across mobile, APIs, and internal networks
- Automated scanners produce shallow results and high false positives
- Lack of contextual awareness, business logic flaws are often missed
- Vulnerabilities are reported without assessing exploitability
- No risk mapping aligned with compliance frameworks
- Heavy dependence on human effort for technical validation

Real attackers don't use checklists; they use creativity, logic, and chaining techniques to break through.

This is precisely where SelfHack AI excels.

.02
SelfHack AI's Solution

Revolutionizing penetration testing with autonomous, intelligent agents.

Traditional penetration testing relies heavily on human effort, limited resources, and slow processes. At best, it offers a snapshot of vulnerabilities at a single moment in time. At worst, it overlooks subtle vulnerabilities hidden deep within system architectures. SelfHack AI completely reimagines this process.

SelfHack AI introduces an AI Pentest Agent architecture that moves beyond traditional scanning. These agents actively think, map attack flows, and evaluate vulnerabilities in context. Our system doesn't just list findings, it evaluates the severity and impact of chained vulnerabilities within your organization's unique environment. SelfHack AI acts not as a "scanner", but as a contextual attack engineer.

With AI agents:

- Recon the environment and analyze logical flows
- Prioritize vulnerabilities based on contextual risk
- Actively exploit and validate findings
- Chain multiple issues to form attack paths
- Report findings in both technical and strategic formats

This model introduces key advantages:

- Eliminates testing gaps with continuously running agents
- Learns from every test and detects pattern similarities
- Low false positive rate, all findings are practically validated
- All results include contextual, technical, and compliance perspectives
- Supports both offensive (Red Team) and defensive (Blue Team) operations

Bottom line: SelfHack AI doesn't replace tools, it scales human expertise. The result is a sustainable, deep, and meaningful penetration testing capability for modern security teams.

Context-Aware Testing

SelfHack AI's core advantage lies in its ability to provide context-aware testing. Conventional scanners often generate findings without considering how those vulnerabilities relate to one another in a live environment. In contrast, SelfHack AI understands attack flows, identifies interconnected weaknesses, and simulates potential attack paths based on real-world adversarial behavior.

This means that SelfHack AI doesn't just tell you that a vulnerability exists, it helps you understand how that vulnerability could be exploited in combination with others, identifying risk in the context of your specific environment. It allows organizations to view threats as attackers would, mapping out the logical progression of exploits from one weakness to the next.

For example, rather than simply identifying a cross-site scripting (XSS) vulnerability, SelfHack AI will simulate how that XSS vulnerability could be used to hijack sessions and ultimately bypass authentication systems. providing a more comprehensive picture of potential threats.

Real-Time, Continuous Testing

Cybersecurity threats are no longer static. With SelfHack AI, penetration testing becomes a continuous, real-time process that is not bound by scheduled intervals. Once deployed, AI agents continuously monitor the security posture of an organization’s infrastructure, ensuring that new vulnerabilities are detected as soon as they emerge.
This ongoing testing ensures that the defense posture is always current and resilient to new, adaptive attacks. It also means that businesses don’t have to wait for a quarterly penetration test to detect a critical vulnerability, it is identified and acted upon immediately.

Autonomous Validation and Exploitation

One of the most powerful features of SelfHack AI is its autonomous validation of vulnerabilities. Rather than relying on human analysts to manually validate each finding, SelfHack AI actively exploits vulnerabilities in a safe and controlled manner to validate their severity and potential impact. This ensures that every finding is meaningful and practical, eliminating the need for time-consuming, error-prone manual validation.
The AI agents mimic the tactics of sophisticated adversaries, using multiple techniques to exploit vulnerabilities and test for real-world attack scenarios. This process results in fewer false positives and ensures that only actionable, verified findings are reported.

Prioritized Risk Reporting

Instead of overwhelming teams with hundreds of unorganized findings, SelfHack AI prioritizes vulnerabilities based on their potential impact to your business. It considers factors such as:

- Exploitable attack paths: How easily an attacker could move through your system.
- Business-critical assets: Which systems or data are most valuable or essential to your operations.
- Risk to compliance: How vulnerabilities could affect your ability to meet regulatory standards.

SelfHack AI automatically assigns severity levels to each vulnerability based on the likelihood of exploitation and potential damage, streamlining the remediation process.

Platform Overview

A self-improving, autonomous cybersecurity engine.

SelfHack AI is built as an intelligent, modular, and scalable platform that enables businesses to detect and understand cyber risks in real time, without relying on manual testing or external consultants. It combines advanced artificial intelligence, autonomous agents, and deep system analysis to deliver continuous penetration testing that adapts to your infrastructure.

At the heart of the platform are our AI Pentest Agents, virtual entities that autonomously explore digital environments, simulate attack chains, exploit vulnerabilities, and learn from each interaction. These agents operate within a secure sandboxed framework to ensure safe execution, while producing actionable insights tailored to the organization's actual risk exposure.

Key Components of the SelfHack AI Platform:

Agent Engine
An advanced orchestration layer that coordinates how AI agents discover, test, and attack environments in real time. The engine dynamically adjusts tactics based on the systems it encounters, mimicking how real-world attackers pivot and adapt.

Target Mapping Module
This module identifies assets, entry points, exposed services, and system hierarchies, forming a live map of the attack surface. It continuously updates as changes are made to the infrastructure.

Vulnerability Chain Simulator
Instead of treating vulnerabilities as isolated issues, this feature simulates chained attack paths to identify how a minor weakness could be leveraged in combination with others to compromise critical systems.

Learning & Adaptation Core
Every interaction improves the system. With each test, the AI agents learn more about the environment, refine their methods, and enhance the depth of future analyses. Over time, the platform builds a knowledge graph of threat behaviors unique to each organization.

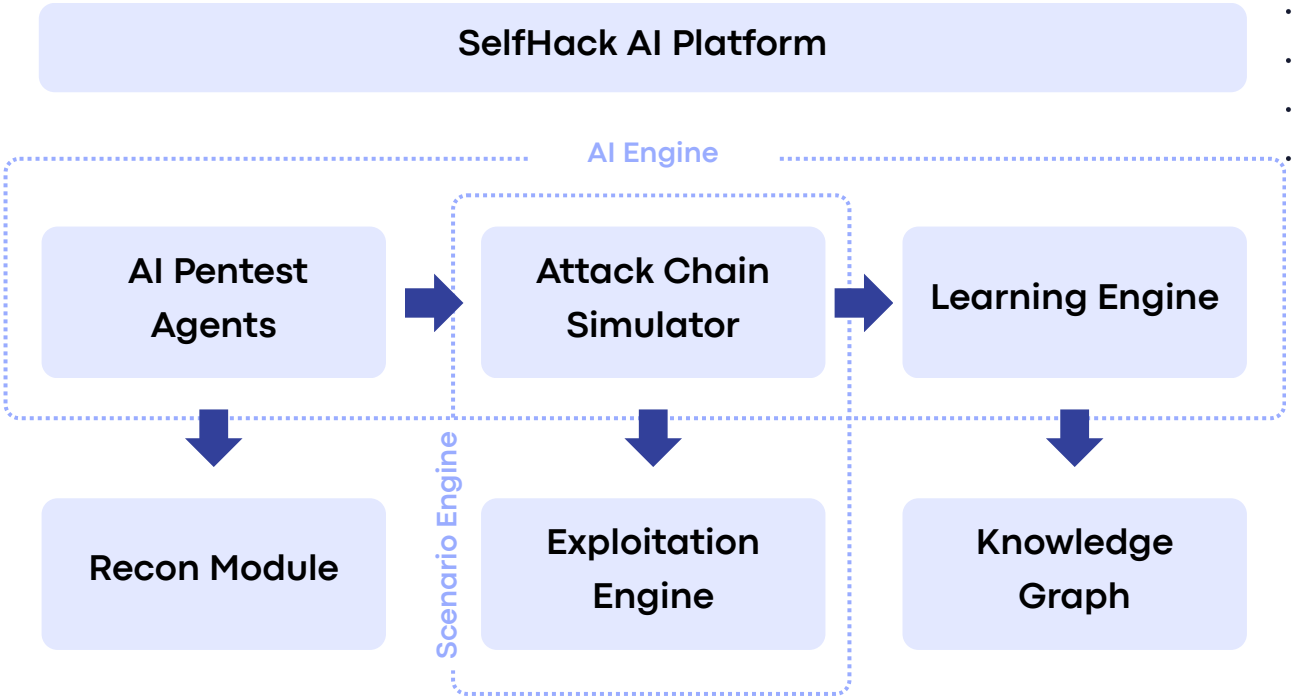
Prioritization & Reporting Dashboard
Findings are surfaced in a clean, business-friendly interface. Critical risks are highlighted based on business impact, exploitability, and compliance relevance, making it easy for technical and non-technical stakeholders to take action.

Integration & API Layer
SelfHack AI integrates seamlessly with existing DevOps pipelines, security monitoring tools, and ticketing systems. Whether you're working with GitHub, Jira, AWS, or Kubernetes, the platform fits into your existing ecosystem to support agile remediation.

SelfHack AI is cloud-native and infrastructure-agnostic. Whether you're a growing startup or a large enterprise with a complex hybrid cloud environment, the platform scales with your needs, from a single microservice to a full-scale enterprise network. Our intuitive deployment model allows customers to get started within minutes. Once deployed, AI agents begin scanning and testing automatically, with no need for extensive configuration or training. The result? Faster, smarter, and continuous penetration testing, without the bottlenecks.

Technical Architecture

A self-improving, autonomous cybersecurity engine.



Reporting & Dashboard

- Technical Reports
- Executive Summaries
- Compliance Mapping

Integration Layer

- CI/CD Tools (e.g., Jenkins, GitLab)
- Communication Tools (e.g., Slack, Teams)
- Issue Tracking (e.g., Jira)

AI Engine: NLP, symbolic reasoning, and contextual attack logic
Knowledge Graph: Continuously learning, relationship-aware intelligence
Scenario Engine: MITRE ATT&CK-aligned chained attack generation
Dashboard: Test management, scheduling, and compliance mapping
Integrations: Jira, SIEM, SOAR, Slack, Teams
Data Security & Isolation: TLS 1.3 for transmission, AES-256-GCM for storage. Built on Zero Trust Architecture and Defense-in-Depth principles. Each tenant is physically and logically isolated. IAM policies are fully customizable, and all access events are archived for legal audits.

AI Components and Methodology

Understanding the Brains Behind Autonomous Penetration Testing

Modern cyber threats require more than just automation, they demand intelligence. SelfHack AI's architecture is built around modular AI components that work together to mimic the reasoning, adaptability, and creativity of a skilled human pentester. The methodology behind our autonomous agents is designed for precision, learning, and scalability.

AI Engine: Emulating Human Reasoning at Scale

At the heart of SelfHack AI lies a powerful engine built on a combination of Natural Language Processing (NLP), symbolic reasoning, and contextual logic modeling. This engine doesn't just parse code or scan endpoints, it understands the logic of an application, identifies flawed flows, and reasons about how to exploit them.

- NLP helps the agent analyze source code, API documentation, and system outputs in human-readable formats.
- Symbolic reasoning allows the system to build logical models of workflows and infer security implications.
- Contextual logic modeling enables adaptive decision-making during live test sessions.

Knowledge Graph: Continuous Learning and Situational Awareness

Every test the system performs feeds back into a dynamic, relationship-aware knowledge graph. This structure captures the relationships between assets, user roles, vulnerabilities, system behaviors, and previous attack patterns. The result is a constantly evolving map of the organization's threat landscape.

- Detects recurring patterns in misconfigurations
- Builds asset-behavior relationships for each environment
- Supports proactive identification of logic flaws beyond CVEs

Scenario Engine: Chained, Real-World Attack Simulation

Instead of reporting isolated bugs, SelfHack AI's Scenario Engine builds chained attack sequences based on real-world adversarial behavior. Aligned with frameworks like MITRE ATT&CK, this engine evaluates what a determined attacker would do after gaining a foothold.

- Generates attack graphs tailored to each environment
- Models lateral movement, privilege escalation, and domain-wide compromise
- Prioritizes findings based on potential business impact, not just CVSS scores

Methodology: Redefining How Testing is Done

SelfHack AI operates under a "Red Team-in-a-Box" philosophy. It blends reconnaissance, vulnerability analysis, logical reasoning, and exploitation in a loop that refines itself with each iteration.

1. Reconnaissance: Map external and internal assets, behavior flows, user roles
2. Analysis: Detect weak points, logic flaws, insecure configurations
3. Exploitation: Attempt real-world attacks to validate findings
4. Chaining: Link multiple issues to simulate breach paths
5. Reporting: Document technical and strategic risks, with remediations

This methodology ensures high accuracy, continuous improvement, and deep contextual understanding, qualities typically reserved for top-tier human red teams.

Scope & Testing Depth

Understanding the Brains Behind Autonomous Penetration Testing

What makes SelfHack AI unique is not just the identification of known issues, it's the ability to conduct deep, contextual, and chained attack simulations across all layers of the environment. Each agent acts like a real-world attacker, simulating not only current threats but forecasting potential attack paths. Even non-CVE vulnerabilities are detected and evaluated based on behavioral insights.

Web Application Pentesting

- Full coverage of OWASP Top 10 + OWASP API vulnerabilities
- Context-aware detection of business logic flaws, privilege escalations
- Deep client-side analysis in modern frameworks (React, Angular)
- Chaining attacks: e.g., XSS → session hijacking → auth bypass

External Network Pentesting

- AI-enhanced Nmap fingerprinting & contextual service analysis
- CVE exploitation, SSL/TLS misconfigurations, default credentials
- Subdomain takeover, CDN abuse, metadata leaks, port misconfigs

Mobile Application Pentesting

- iOS & Android: static + dynamic analysis, binary reverse engineering, traffic inspection
- Detection of insecure storage, hardcoded secrets, debug leaks
- Jailbreak/root detection, SSL pinning bypass, dangerous permissions

Internal Network Pentesting

- Segment-based network discovery via VPN and lateral movement
- Pass-the-hash, Kerberoasting, rogue services, privilege chains
- Active Directory: user enumeration, domain escalation, RCE → DC takeover

API Pentesting

- REST, GraphQL, SOAP: endpoint discovery + parameter fuzzing
- BOLA, mass assignment, token manipulation, improper rate limiting
- Swagger/OpenAPI-based test generation and logic flaw detection

Social Engineering

- Realistic spear phishing, QR phishing, and credential harvesting
- Behavior-driven targeting and MFA bypass simulation
- Voice phishing (vishing) and SIM swap simulation

Every test goes beyond isolated issues, it measures the true resilience of your organization through real-world attack paths. The goal is not only to "find" vulnerabilities but to contextualize what they could become.

Key Features

- Actionable solutions at code and server levels
- End-to-end test automation (0-touch model)
- AI-driven contextual analysis and attack chain modeling
- Executive-friendly outputs: compliance, impact, risk scoring
- Technical team-ready reports: PoC, payloads, logs, code/config fixes

Use Cases

t its core, SelfHack AI is an intelligent pentest agent. It behaves like a security expert: it analyzes the target, finds weaknesses, chains them into meaningful attacks, and delivers context-rich reports. Real-world use cases include:

- Autonomous Red Team Operations (Red Team-in-a-Box)
- Chained Vulnerability Simulations
- Live Attack Surface Monitoring
- Threat-Informed Testing (MITRE ATT&CK-Aligned)
- Continuous Penetration Testing (CPT)
- Compliance-Ready Output
- Hybrid Use with Red and Blue Teams

Vision

SelfHack AI aims to orchestrate AI-powered Red, Blue, and Purple Team agents working in harmony.

- Autonomous decision-making agents
- Real-time AI SOC capable of incident response
- AI-first sustainable cybersecurity defense frameworks

Security & Compliance

In cybersecurity, trust is not granted, it's earned through transparency, resilience, and adherence to rigorous standards. SelfHack AI has been engineered from day one with security, data protection, and compliance as foundational principles.

Zero Trust Architecture

Our infrastructure follows a Zero Trust Architecture (ZTA) model. Every interaction, from internal agent actions to API requests, undergoes authentication, authorization, and continuous validation. This approach eliminates implicit trust, enforcing strict segmentation across users, data, and processes.

- **Micro-segmentation:** Logical separation of data, infrastructure, and workloads per tenant
- **Per-request validation:** Continuous authentication and authorization using modern IAM policies
- **Least privilege access:** Every system component only receives the exact permissions it needs

Data Protection & Encryption

All sensitive data in SelfHack AI is protected by multi-layer encryption, both in transit and at rest:

- TLS 1.3 for all network communications
- AES-256-GCM for encrypted storage
- Key management via secure, auditable vaults
- Immutable logging of access and changes for audit trails

This ensures compliance with GDPR, HIPAA, and other global data protection mandates.

Tenant Isolation

Each customer environment is physically and logically isolated, with no shared compute, memory, or database instances. This guarantees that even in a multi-tenant setup, data from one client cannot be accessed or influenced by another.

- Dedicated execution environments per tenant
- Isolated vulnerability and test data storage
- Scoped IAM and RBAC policies tailored to each organization

Compliance-Ready & Audit-Friendly

SelfHack AI was built for organizations operating in regulated sectors such as finance, healthcare, critical infrastructure, and SaaS. We align with common security and compliance standards to make adoption seamless:

- SOC 2 alignment in logging, access controls, and operations
- GDPR-ready, with data processing agreements and data subject rights management

- MITRE ATT&CK framework used in reporting and testing methodology

Audit logging and archiving for legal, forensic, and governance needs

Secure Development Lifecycle (SDL)

Our development and deployment processes follow a secure software development lifecycle, with continuous code review, automated security testing, dependency scanning, and post-deployment validation.

- Static and dynamic analysis on each build
- Secure CI/CD pipelines with signed artifacts
- Real-time vulnerability monitoring and remediation workflows

.03
Use Cases

SelfHack AI's autonomous agents adapt to different layers of your digital infrastructure, not only recognizing surface-level issues, but contextualizing each vulnerability within its architectural, business, and regulatory context.

Web Applications

- **Beyond OWASP Top 10** : AI agents identify and exploit weaknesses in authentication workflows, access controls, session handling, and client-side logic across modern stacks like React, Angular, and Vue.
- **Stateful Attack Modeling:** Tracks user journeys across pages, cookies, and dynamic parameters to construct realistic, multi-step attacks (e.g., auth bypass → cart manipulation → role elevation).
- **Regulation-Aware Findings:** Each vulnerability includes mapping to applicable standards such as ISO 27001 A.12.6.1, PCI-DSS 6.5.x, or GDPR Article 32, with risk grading based on business-critical functionality.
- **CI/CD Integrations:** Web application tests can be scheduled post-deploy or post-merge via integrations with GitHub Actions, GitLab CI, or Bitbucket Pipelines.

Mobile Applications

- **Full-Stack Mobile Analysis:** iOS/Android applications are decompiled, analyzed, and executed in sandboxed environments. Agents inspect runtime behaviors, insecure storage, exposed debug APIs, and client-server interactions.
- **Code-Level Vulnerability Discovery:** Includes detection of hardcoded secrets, improper SSL pinning, root/jailbreak indicators, and local data leaks (SharedPreferences, SQLite, plist, etc.).
- **Contextual Agent Logic:** AI agents correlate app behavior with backend APIs to simulate real-world attack vectors (e.g., manipulating API calls after modifying local state or capturing JWTs from traffic).
- **DevOps Integrations:** SelfHack AI integrates with mobile CI/CD pipelines (e.g., Fastlane, Bitrise) and stores build artifacts securely during testing. Results include full PoC flows and remediation code snippets.

APIs (REST, GraphQL, SOAP)

- **Dynamic Endpoint Discovery:** Agents crawl OpenAPI, Swagger, GraphQL introspection, and hidden request patterns to fully map available attack surface, even for undocumented endpoints.
- **Parameter Fuzzing + Authentication Context:** Identifies weak token validation, BOLA (Broken Object Level Authorization), mass assignment, enum abuse, and bypasses for API keys, JWTs, or OAuth flows.
- **Stateful Flow Reproduction:** Vulnerabilities are not flagged unless the full exploitation flow can be replicated and contextualized (e.g., "unauthorized user accesses another user's invoice via direct ID enumeration").
- **Compliance-Tagged Reporting:** API findings include ISO 27001 Annex A mappings (A.9.1.2, A.13.2.1) or NIST SP 800-53 (AC, SC controls), with automated risk grading by SelfHack AI's compliance matrix.
- **Webhook-Based Testing Triggers:** Run targeted tests when a new API is deployed or updated, using a single webhook integrated into your API gateway or API management platform.

External Network

- **AI-Enhanced Recon & Fingerprinting:** Self-learning agents identify open services, outdated software, exposed portals, subdomain takeovers, SSL misconfigs, and unauthenticated panels.
- **Exploitation-Driven Risk Prioritization:** Only vulnerabilities that are exploitable (not theoretical) and accessible from the outside world are flagged, each tied to possible compliance violation vectors.
- **Asset Inventory Drift Detection:** Agents automatically detect new subdomains, DNS changes, open ports, or cloud asset exposure, ideal for organizations with complex, rapidly changing environments.
- **Enrichment via Integrations:** Integrate findings with external ASM tools, asset inventories, CMDBs, and vulnerability management platforms (e.g., ServiceNow, Wiz, Axonius) for live correlation.

Internal Network

- **Segment-Aware Reconnaissance:** Via VPN or direct agent deployment, SelfHack AI maps internal infrastructure by scanning for service banners, authentication exposures, and domain trust relationships.
- **Credential Abuse & Lateral Movement:** Simulates realistic privilege chains (e.g., SMB share access → password reuse → AD escalation), validates exposure without destructive actions.
- **Active Directory Targeting:** Performs user/group enumeration, Kerberoasting, AS-REP Roasting, and password spraying, with attack graphs showing how initial access can become full domain control.
- **Zero Trust Validation:** Evaluate segmentation, MFA enforcement, default account exposure, and shadow IT risks, aligned with compliance controls like ISO A.13.1.1 or NIST AC-4.
- **Internal Integration Capabilities:** Test results can be pushed directly into internal SIEM/EDR pipelines or asset governance dashboards, tagging each finding with hostnames, AD objects, or user IDs.

APIs (REST, GraphQL, SOAP)

- **Dynamic Endpoint Discovery:** Agents crawl OpenAPI, Swagger, GraphQL introspection, and hidden request patterns to fully map available attack surface, even for undocumented endpoints.
- **Parameter Fuzzing + Authentication Context:** Identifies weak token validation, BOLA (Broken Object Level Authorization), mass assignment, enum abuse, and bypasses for API keys, JWTs, or OAuth flows.
- **Stateful Flow Reproduction:** Vulnerabilities are not flagged unless the full exploitation flow can be replicated and contextualized (e.g., "unauthorized user accesses another user's invoice via direct ID enumeration").
- **Compliance-Tagged Reporting:** API findings include ISO 27001 Annex A mappings (A.9.1.2, A.13.2.1) or NIST SP 800-53 (AC, SC controls), with automated risk grading by SelfHack AI's compliance matrix.
- **Webhook-Based Testing Triggers:** Run targeted tests when a new API is deployed or updated, using a single webhook integrated into your API gateway or API management platform.

Sector Use Cases: Telecommunications

Use Case: Telco operators deploy SelfHack AI agents to continuously test customer-facing portals, B2B provisioning APIs, and internal CRM flows for misuse and logic-level issues.

Real-World Need:

- SIM provisioning APIs are often exposed to resellers → risk of over-provisioning or account hijack
- Internal customer support tools sometimes bypass MFA due to legacy exceptions

What They Do:

- Agents are deployed in test environments or via sandboxed sessions on staging servers

- Tests are scheduled weekly and integrated into Jira for each region's DevSecOps team
- CVEs are de-prioritized; instead, agents look for logic chains like "porting a number without user confirmation"

Regulation Alignment:

- GDPR (data access), ISO 27011 (Telco InfoSec), country-specific telco laws (e.g., KVKK, CCPA)

Sector Use Cases: Energy / Utilities / Smart Infrastructure

Use Case: Energy providers use SelfHack AI to map their externally facing digital assets (admin panels, OT portals, cloud services) and identify weak links between IT and OT networks.

Real-World Need:

- Many plants have exposed RDP or VNC due to remote maintenance contracts
- Cloud dashboards used by field engineers are often misconfigured or outdated

What They Do:

- Agents run monthly perimeter scans and simulate logic flows in role-specific dashboards (e.g., field engineer vs admin)
- Findings are sent to a centralized dashboard; local facility managers receive only their filtered asset risks
- No SCADA/ICS device is touched, only the attack paths toward them are assessed

Regulation Alignment:

- IEC 62443 (OT), NIS2, ISO 27019 (energy vertical)

Sector Use Cases: Financial Institutions & Fintech

Use Case: Banks and fintechs deploy SelfHack AI for weekly logic-level testing on customer-facing apps, money movement features, and KYC workflows.

Real-World Need:

- Mobile apps allow changing withdrawal limits without sufficient secondary auth
- Legacy investment portals use SSO tokens across subdomains, can be reused for privilege escalation

What They Do:

- A separate AI agent profile is created for each product line: banking, credit, investment
- Staging environment is mirrored weekly; agents test the full login → transaction → logout flow
- Each finding includes a regulatory tag (e.g., PCI DSS 6.5.2 or PSD2 Article 97)

Integration Examples:

- Reports are pushed to GRC tools like RSA Archer or ServiceNow GRC for workflow
- PoCs are optionally sent to internal QA engineers for retesting prior to release

Sector Use Cases: Conglomerates & Holdings

Use Case: Large enterprise groups deploy one SelfHack AI instance per subsidiary and coordinate risk aggregation at the group level.

Real-World Need:

- Each subsidiary uses different authentication, RBAC, and CI/CD stacks, causing inconsistent defenses
- Group CISO lacks visibility into weak links introduced by B2B integrations or SSO bridges

What They Do:

- Each subsidiary receives its own isolated tenant and AI agent behavior configuration
- Monthly reports are automatically compared across entities to find "cross-subsidiary misalignment"
- Shared vulnerabilities (e.g., same outdated JS lib used across 3 subsidiaries) are flagged as systemic risk

Integration Examples:

- Slack/Teams alerts notify InfoSec when an issue is seen in more than one entity
- A central dashboard visualizes which subsidiaries pass minimum security SLAs

Sector Use Cases: Healthcare & Pharma

Use Case: Clinical platforms and healthcare providers use SelfHack AI to test patient portals, trial data platforms, and doctor-facing mobile apps.

Real-World Need:

- Patient data often accessible with session reuse
- Admin portals have hidden debug features left active in production

What They Do:

- Mobile and web agents are rotated between test and pre-prod environments
- Session hijacking and IDOR attempts are modeled, but agents are scoped to non-production databases
- Findings include not only tech detail but potential HIPAA Security Rule implications

Integration Examples:

- Integration with internal ticketing (e.g., Jira or Asana)
- Findings flagged with severity + risk to PHI exposure

Future Roadmap

01

Fraud Risk Analysis Agent (Q1 2026)

To identify abuse potential in business processes, financial flows, and user roles that may enable internal or external fraud scenarios.

02

Intelligence Risk Detection Agent (Q2 2026)

To identify areas within the application or infrastructure that could leak metadata, internal logic, or operational secrets, useful for social engineering, industrial espionage, or APT targeting.

03

Trust Surface Mapping (Q3 2026)

To map systems, roles, and components not just as "assets" but as trust boundaries, highlighting where implicit trust creates attack potential.

+ + +

Expected Outputs

- “Abuse Chain Reports” showing how a user can turn a legitimate function into a vector for fraud
- Fraud risk scores, tagged with financial and reputational impact
- ISO 27001 A.15.2.1 / PCI DSS / SOX relevance included per case
- “Intelligence Footprint” reports that outline how much an external attacker can learn before any exploit
- Pretext vectors (used in phishing, impersonation, social engineering)
- Strategic risk flags (e.g., leaked contract references, exposed internal project codenames)
- Trust breakdown reports, showing lateral movement paths between roles or services
- Suggested zero-trust remediation steps
- Tagging for NIST 800-207, ISO A.9 & A.13 alignment

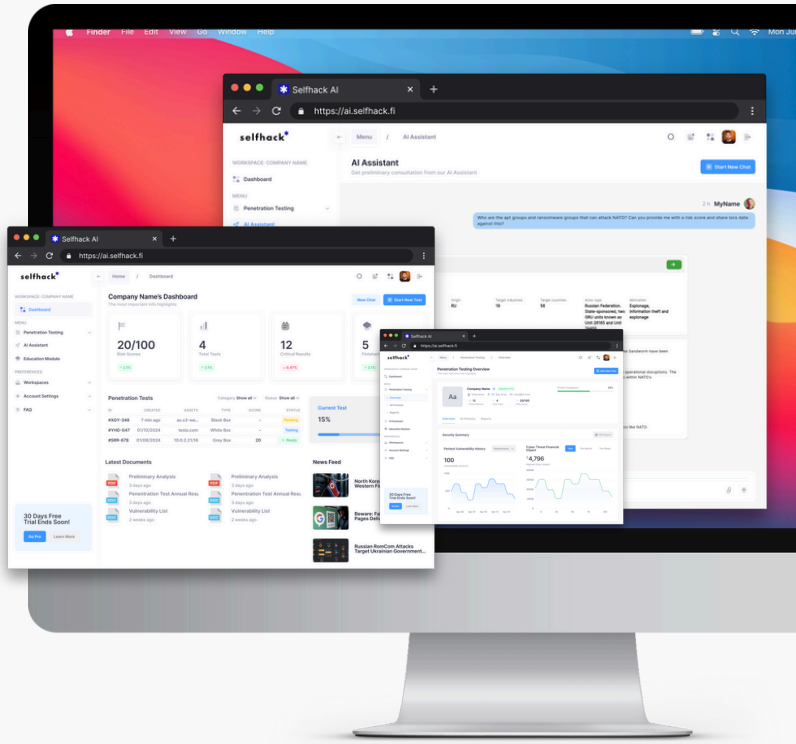
.04
Appendix

References & Supporting Sources

1. IBM Security. "Cost of a Data Breach Report 2023." IBM, in collaboration with Ponemon Institute. https://www.ibm.com/reports/data-breach
2. Verizon. 2023 Data Breach Investigations Report (DBIR). https://www.verizon.com/business/resources/reports/dbir/
3. MITRE ATT&CK Framework. https://attack.mitre.org/

Benchmark Scope

The SelfHack AI benchmark encompasses over 10, 200 real-world vulnerabilities collected from:



- CVE
- NVD
- HackerOne Bug Bounty Open Source Reports
- ExploitDB
- Bugcrowd Bug Bounty Open Source Reports
- OWASP Top 10
- Mitre Attack
- All Clear Web and Dark Web Source etc.

Platform-Specific Success Rates:

99% Web Applications

96% Mobile Applications

98% APIs

Detailed Breakdown

Category	True Positive	False Positive	Exploit Success Rate	Automation
Web	99%	1%	95%	98%
Mobile	96%	4%	93%	96%
API	98%	2%	94%	97%

+ + +

SelfHack AI's superior performance is underpinned by five key features:

- **Pattern Recognition**: Understands new threat types and proactively responds to attacks.
- **Dynamic Exploit Simulation**: Automatically adapts payloads and attack methods to the target.
- **Root Cause Analysis**: Identifies and reports vulnerabilities based on root causes rather than symptoms.
- **Automated Remediation**: Offers code and server-level recommendations, supporting all major programming languages.
- **Continuous Learning**: Learns from each test, building resilience against future threats.

Real-World Use Cases

Cyber threats don’t wait, neither should your defenses.

Companies across industries trust Self Hack AI* to uncover hidden vulnerabilities, automate remediation, and elevate their security posture. From cloud security to e-commerce platforms, see how AI-driven penetration testing is transforming cybersecurity.

CloudFirm CEO – Cloud Service Provider Security

“Manual testing could not keep pace with the rapid threats we encountered on our cloud infrastructure. SelfHack AI automatically discovered unnoticed vulnerabilities within our cloud servers, providing configuration recommendations that enhanced our overall security posture and operational compliance.”

E-Commerce Platform CTO – Online Platform Security

“SelfHack AI identified critical SQL Injection and IDOR vulnerabilities etc. within seconds on our e-commerce platform and provided immediate remediation solutions. This significantly boosted customer trust and elevated our security practices to international standards.”

MSSP Security Director – Managed Security Services

“ We performed extensive penetration tests in Belgium, France, and Germany for our diverse customer base, delivering customized, localized, and comprehensive reports in multiple languages, thanks to SelfHack AI. This automated and intelligent solution allowed us to effectively address each customer’s unique security needs and significantly enhance the quality of our MSSP services.”

**These are real scenarios from our customers who have experienced the power of SelfHack AI * firsthand. Due to NDAs, we can’t disclose their identities, but their stories speak for themselves, showcasing how AI-driven penetration testing is transforming cybersecurity across industries.*

Competitor Comparison

Tool	True Positive	False Positive	Exploit Success Rate	Automation
SelfHack AI	99%	1%	95%	98%
Tool #1	85%	15%	70%	60%
Tool #2	88%	12%	75%	65%
Tool #3	88%	12%	70%	60%
Tool #4	84%	16%	65%	55%
Tool #5	91%	9%	75%	40%

To ensure fairness and maintain professionalism, we've anonymized competitor names while providing descriptive labels that reflect their functionality. This comparison is based on publicly available data and internal testing, allowing readers to focus on performance insights rather than brand names. Our goal is to highlight how different security solutions measure up without bias. To Learn More Contact Us

1. *Leading Web Security Scanner – A commercial tool widely used for detecting vulnerabilities in web applications, particularly known for its automated scans and ease of use.*
2. *Cloud-Based Security Platform – A security suite offering website protection, malware scanning, and penetration testing, commonly used by SMBs and e-commerce businesses.*
3. *Enterprise Vulnerability Scanner – A well-known vulnerability assessment tool designed for enterprise environments, focusing on identifying security gaps in networks and systems.*
4. *Open-Source Security Scanner – A free and open-source vulnerability scanning framework, often used by security professionals for detecting weaknesses in IT infrastructure.*
5. *Advanced Web Security Testing Tool – A powerful tool favored by penetration testers for manual and semi-automated web security testing, particularly for complex vulnerabilities.*

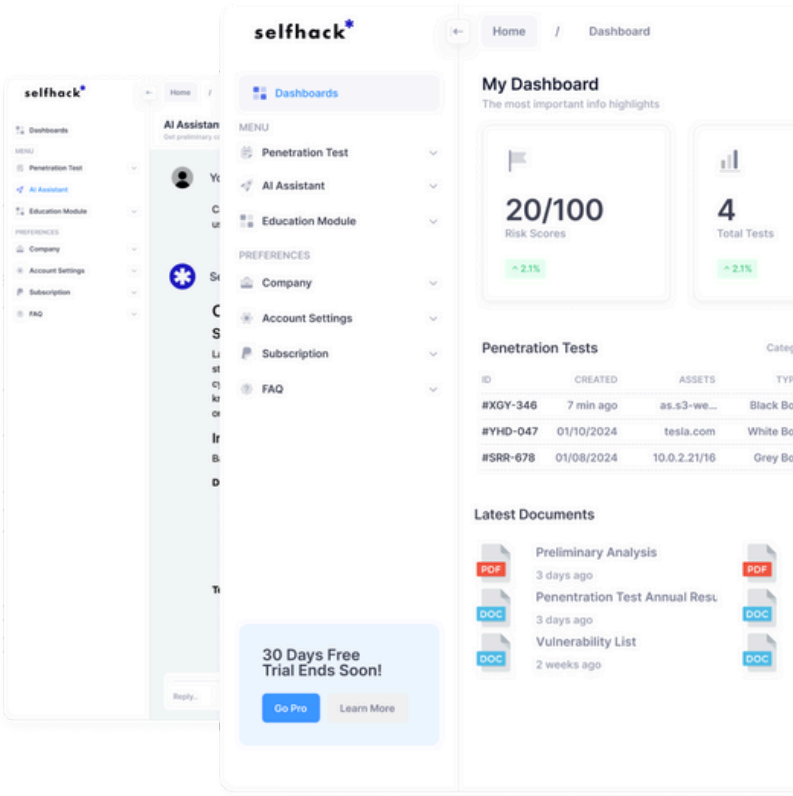
Conclusion: Secure the Future with Selfhack AI*

In today's volatile threat landscape, reactive security measures are no longer enough. Threat actors evolve rapidly, and only proactive defense strategies can ensure your organization stays one step ahead.

Why Selfhack AI*?

Selfhack AI revolutionizes cybersecurity by automating and streamlining periodic penetration testing, offering you:

- **Proactive Security:** Identify vulnerabilities before they're exploited.
- **Autonomous Testing:** Save time and resources with automated, AI-driven testing tailored to your systems.
- **Comprehensive Insights:** Receive detailed reports, actionable recommendations, and targeted training modules.
- **Continuous Protection:** Stay protected against emerging threats with ongoing testing and risk scoring.



How It Works:

Selfhack AI empowers your security teams by performing:

1. Automated penetration tests across Web, Mobile, API, VoIP, and more.
2. AI-generated risk assessments with statistics and mitigation strategies.
3. Education modules to strengthen your team's expertise.

Take Control Today with Selfhack AI!

Don't wait for threats to become crises. By integrating Selfhack AI into your cybersecurity strategy, you'll strengthen your defenses, meet compliance requirements, and build trust with stakeholders. Visit www.selfhack.fi to learn more about Selfhack AI and take the first step towards autonomous security today!

selfhack*

Contact Us*

Explore the Future of Cybersecurity with Selfhack AI*!

We appreciate your interest in Selfhack AI. As a special offer, the first 15 companies to contact us using the code ‘SAFE25DEMO’ will receive a free demo of our platform!

Scan the QR code to get in touch, and our team will guide you through the next steps to secure your business. Don’t miss this opportunity to experience proactive cybersecurity at its best!

ADDRESS

Maria 01
Helsinki, Finland

CONTACT US

info@selfhack.fi
www.selfhack.fi