OWASP | GenAI SECURITY PROJECT

AI SECURITY SOLUTIONS INITIATIVE

Q2 2026

AI Security Solutions Landscape

For AI and Agentic Red Teaming

https://genai.owasp.org/ai-security-solutions-landscape/

OWASP | GenAI SECURITY PROJECT

AI SECURITY SOLUTIONS INITIATIVE

Q2 2026

AI Security Solutions Landscape

For AI and Agentic Red Teaming

As organizations increasingly deploy generative AI and autonomous agents into business-critical workflows, traditional application security practices are no longer sufficient. AI systems introduce new classes of risk including prompt injection, model misuse, agent privilege escalation, data poisoning, hallucinations, and emergent behaviors that evolve continuously throughout the AI adoption lifecycle. Gen AI and Agentic Red Teaming provides a structured, lifecycle-wide approach to identifying, measuring, mitigating, and governing these risks through coordinated adversarial testing, defensive validation, and continuous feedback loops.

https://genai.owasp.org/ai-security-solutions-landscape/

OWASP | GenAI SECURITY PROJECT

CHEAT SHEET genai.owasp.org

Red Teaming Landscape – Q2 2026

https://genai.owasp.org/ai-security-solutions-landscape/

Operate

ADVERSA, TROJ.AI, fiddler, Straiker, paloalto PRISMA AIRS, ALICE, NOMA, HIDDENLAYER, highflame, TREND MICRO, opsin, zscaler, Edictum, NRI SECURE Deep AI Red Team, exabeam, XecGuard, LASSO, VULCAN, tenuo, f5

Deploy

NOMA, ADVERSA, HIDDENLAYER, fiddler, Dynamo AI, opsin, zscaler, highflame, LASSO, Straiker, ALICE, paloalto PRISMA AIRS, XecGuard, NRI SECURE Deep AI Red Team, TREND MICRO, humanbound, f5, tenuo

Release

ALICE, LASSO, highflame, paloalto PRISMA AIRS, NRI SECURE Deep AI Red Team, zscaler, ADVERSA, HIDDENLAYER, NOMA

Test & Evaluate

highflame, HiveTrace.red, zenity, AI War Games, ALICE, FUJITSU, VULCAN, NRI SECURE Deep AI Red Team, HIDDENLAYER, LASSO, NOMA, humanbound, Surelio.ai, ADVERSA, TROJ.AI, SISA AI PRISM, TREND MICRO, zscaler, tenuo, fiddler, f5, Straiker, XecART, XecGuard, paloalto PRISMA AIRS, Avenlis SandStrike, Reinforce Labs, LLM-RTK

Develop & Experiment

LASSO, ADVERSA, NRI SECURE Deep AI Red Team, NOMA, TROJ.AI, highflame, Surelio.ai, paloalto PRISMA AIRS, Straiker, ALICE, zscaler, f5, XecART, fiddler, HIDDENLAYER, AI

Augment, Fine Tune Data

highflame, zscaler, paloalto PRISMA AIRS, Surelio.ai, ALICE, XecGuard, NOMA

Scope & Plan

ADVERSA, paloalto PRISMA AIRS, XecART, TREND MICRO, ALICE, zscaler, NOMA, tenuo, Surelio.ai, TROJ.AI, fiddler, NRI SECURE Deep AI Red Team, Straiker, LASSO

Monitor

NOMA, tenuo, zscaler, highflame, LASSO, fiddler, TREND MICRO, exabeam, ADVERSA, Surelio.ai, NRI SECURE Deep AI Red Team, paloalto PRISMA AIRS, HIDDENLAYER, opsin, f5, Straiker, humanbound

Govern

Surelio.ai, FUJITSU, ADVERSA, highflame, ALICE, zscaler, LASSO, NRI SECURE Deep AI Red Team, fiddler, TREND MICRO, VULCAN, tenuo, TROJ.AI, NOMA, f5, exabeam, HIDDENLAYER, paloalto PRISMA AIRS, Straiker, opsin, humanbound

Open Source | Gold Sponsors | Silver Sponsors

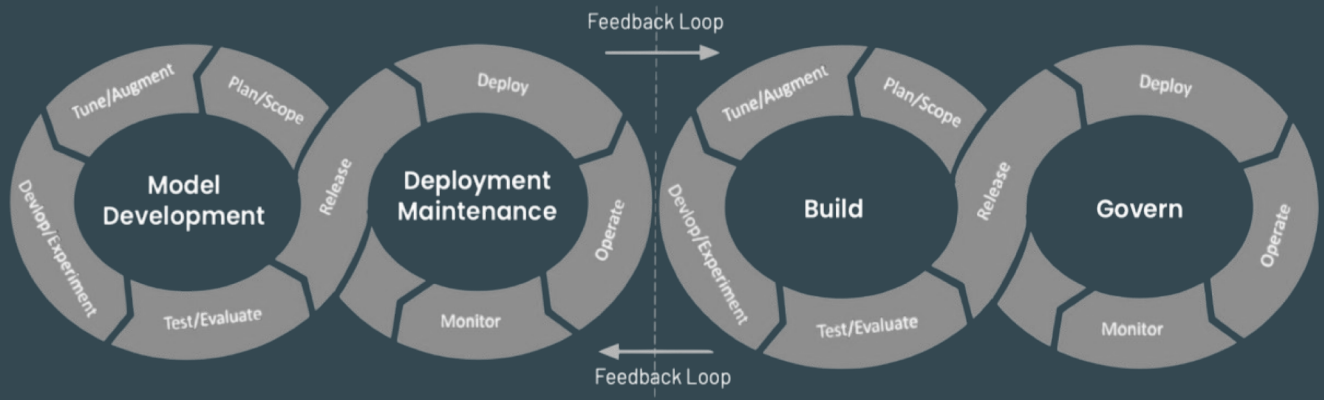
Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming SecOps Framework



The framework organizes capabilities into four categories aligned to security teaming and collaborative security team roles.

Red Teaming (Offense / Attack Simulation)	Focuses on proactively discovering weaknesses in AI models, agents, data pipelines, and integrations by simulating malicious users, compromised agents, and supply-chain attacks.
Blue Teaming (Defense / Detection / Response)	Ensures that guardrails, policies, monitoring, and runtime protections prevent, detect, and respond to AI-specific threats across environments.
Purple Teaming (Continuous Attack–Defense Fusion)	Integrates red and blue activities into closed-loop feedback systems, ensuring adversarial findings directly improve defensive controls, policies, and operational posture.
Shared Capabilities	Foundational capabilities—such as asset inventory, telemetry, provenance, metrics, and audit artifacts—that enable consistency, automation, and governance across all teams.

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

AI and Agentic Red Teaming Landscape – Q2 2026

Scope & Plan

This stage establishes the foundation for AI security by identifying risks, assets, and threat scenarios early. Red teaming focuses on modeling adversaries, mapping attack surfaces, and identifying misuse paths across models, agents, and integrations. Blue teaming builds inventories, risk dashboards, and prioritization frameworks to ensure visibility and governance. Purple teaming connects both by mapping adversarial scenarios to defensive controls, enabling traceability and gap analysis. Together, this phase ensures organizations understand where risks exist, who owns them, and how offensive insights translate into defensive readiness before development begins.

Red/Blue/Purple Teaming Requirements

- Threat-model design aids
- LLM / agent attack-surface mapping
- AI asset inventory
- AI posture dashboards (AI-SPM / AI-TRiSM)
- Risk-scoring boards
- Import red scenarios
- Map red scenarios to blue controls
- Risk taxonomy import/export
- Visual data-flow mapping
- Export of tests as stories

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

AI and Agentic Red Teaming Landscape – Q2 2026

Augment, Fine Tune Data

highflame

zscaler

Surelio.ai

NOMA

ALICE

paloalto NETWORKS PRISMA AIRS

XecGuard

This stage addresses risks in training data and model adaptation. Red teaming stress-tests pipelines using poisoned data, synthetic adversarial inputs, and malicious artifacts to expose weaknesses in learning processes. Blue teaming ensures data integrity through lineage tracking, DLP scanning, and bias/toxicity auditing to prevent harmful or non-compliant outputs. Purple teaming bridges both by replaying adversarial data against defenses and comparing dataset versions to detect regressions. The goal is to ensure that models are trained on trustworthy data, remain robust to manipulation, and that defensive controls evolve alongside changing datasets.

Red/Blue/Purple Teaming Requirements

- Data-poison fuzzing
- Synthetic insert generation
- Malicious model artifacts
- Data lineage & provenance tracking
- DLP scanning
- Bias–toxicity co-auditing
- Replay red mutations through blue filters;
- Corpus diffing
- Bias / PII scorecards
- Signed data packages

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

AI Red Teaming Landscape – Q2 2026

Develop & Experiment

ADVERSA
zscaler
f5
LASSO
Straiker
ALICE
NOMA
AI
XecART
TROJ.AI
NRI SECURE
Deep AI Red Team
Surelio.ai
highflame
fiddler
HIDDENLAYER
paloalto NETWORKS
PRISMA AIRS

During development, teams evaluate model and system behavior before production. Red teaming probes for vulnerabilities such as prompt injection, jailbreaks, and agent logic manipulation. Blue teaming applies traditional and AI-specific security testing (e.g., SAST/DAST, plugin scanning) to code and integrations. Purple teaming enables collaboration through shared sandboxes, signal analysis, and automated remediation workflows when failures are found. This stage ensures vulnerabilities are discovered early, defensive signals are validated, and feedback loops between attackers and defenders improve both model robustness and system security prior to formal testing.

Red/Blue/Purple Teaming Requirements

- Model vulnerability scanning
- Agent-logic corruption testing
- SAST / DAST / IAST scanning
- LLM plugin, tool, and infrastructure scanning
- Interactive sandbox
- Defender signal analysis
- Reasoning-trace capture
- Auto-ticketing for failed tests
- IDE plugins

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Test & Evaluate

highflame
HiveTrace.red
zenity
HIDDENLAYER
LASSO
TREND MICRO
Surelio.ai
humanbound
Avenlis SandStrike
SISA AI PRISM
AI War Games
NRI SECURE Deep AI Red Team
FUJITSU
NOMA
TROJ.AI
XecART
tenuo
zscaler
fiddler
f5
VULCAN
XecGuard
Straiker
Reinforce Labs
ALICE
ADVERSA
LLM-RTK
paloalto NETWORKS
PRISMA AIRS

Open Source Gold Sponsors Silver Sponsors

This phase rigorously validates system resilience. Red teaming executes adversarial test suites, including multi-turn attacks, prompt chaining, and protocol exploits, to uncover emergent failures. Blue teaming verifies that guardrails and policies function correctly under both normal and adversarial conditions. Purple teaming integrates both perspectives through unified test runs, KPI tracking, and success-threshold evaluation, often embedded in CI pipelines. The outcome is a measurable understanding of risk, ensuring that both offensive findings and defensive effectiveness are quantified before release decisions are made.

Red/Blue/Purple Teaming Requirements

- Automated adversarial suites
- Prompt-chaining attacks
- Multi-turn attacks
- Protocol attacks (A2A, MCP)
- RAG-poison scenario runners
- Guardrail conformance testing
- Policy testing & validation
- One-click purple runs
- Metrics exporting (blue KPIs)
- Success-threshold analysis
- Hallucination vs misalignment labeling
- Continuous Integration hooks

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Release

highflame

LASSO

HIDDENLAYER

ALICE

paloalto NETWORKS PRISMA AIRS

zscaler

ADVERSA

NRI SECURE Deep AI Red Team

NOMA

The release stage ensures only secure and validated artifacts reach production. Red teaming simulates supply-chain attacks to test exposure to compromised models, data, or dependencies. Blue teaming enforces secure CI/CD gates, validates provenance, and ensures artifacts are trusted and untampered. Purple teaming adds end-to-end pipeline analysis, risk dashboards, and rollback planning to mitigate release failures. This stage acts as a control checkpoint where offensive insights and defensive safeguards converge to determine whether residual risk is acceptable for deployment.

Red/Blue/Purple Teaming Requirements

- Supply-chain attack simulation
- Secure CI/CD gates
- Signing & provenance validation
- Purple pipeline analysis
- Release-risk dashboards
- Rollback script generation
- AI-BOM / SBOM diffing

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Deploy

highflame · LASSO · XecGuard · tenuo · ADVERSA · HIDDENLAYER · ALICE · Straiker · fiddler · opsin · TREND MICRO · NOMA · f5 · humanbound · Dynamo AI · NRI SECURE Deep AI Red Team · paloalto NETWORKS PRISMA AIRS · zscaler

Deployment focuses on securing runtime environments. Red teaming simulates misuse of tools, privilege escalation, and cross-tenant data leakage in deployed systems. Blue teaming implements runtime protections such as AI firewalls and policy enforcement systems. Purple teaming enhances this with live traffic simulations, shadow policy testing, and protocol spoofing to validate defenses in realistic conditions. This stage ensures that once systems are live, they are hardened against real-world threats and that defensive controls operate effectively under production constraints.

Red/Blue/Purple Teaming Requirements

- Tool-chain / plug-in misuse simulation
- Agent privilege-escalation emulation
- Cross-tenant data exposure testing
- LLM / agent firewall
- Policy management
- Live traffic chaos simulation
- Real-time policy shadow mode
- Protocol spoofing (MCP / A2A)
- Cost-impact tracking

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source · Gold Sponsors · Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Operate

highflame
TROJ.AI
NOMA
LASSO
ADVERSA
TREND MICRO
Straiker
XecGuard
fiddler
tenuo
VULCAN
HIDDENLAYER
ALICE
zscaler
Edictum
NRI SECURE
Deep AI Red Team
opsin
f5
paloalto NETWORKS
PRISMA AIRS
exabeam

In operation, systems face continuous and evolving threats. Red teaming uses autonomous agents, prompt fuzzing, and memory poisoning to actively probe systems over time. Blue teaming monitors runtime behavior, detects anomalies, and responds to incidents using AI-specific security controls. Purple teaming creates closed-loop feedback systems, correlating attacks with alerts, tuning detection rules, and automatically improving guardrails. This stage emphasizes continuous security rather than point-in-time testing, ensuring defenses adapt dynamically as threats evolve.

Red/Blue/Purple Teaming Requirements

- Autonomous red bots
- Continuous prompt fuzzing
- Memory poisoning
- Runtime AI-SPM / AI-WAF
- Anomaly & drift detection
- Trust-boundary alerting
- Closed-loop purple coaching
- Red/blue alert correlation
- Rule tuning
- Auto guardrail patching

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Open Source Gold Sponsors Silver Sponsors

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Monitor

humanbound
ADVERSA
f5
HIDDENLAYER
LASSO
TREND MICRO
opsin
highflame
Surelio.ai
zscaler
Straiker
fiddler
NOMA
NRI SECURE Deep AI Red Team
tenuo
paloalto NETWORKS PRISMA AIRS
exabeam

Open Source | Gold Sponsors | Silver Sponsors

Monitoring provides visibility into system health and security posture. Red teaming contributes by generating synthetic malicious users and rogue agents to test detection capabilities. Blue teaming collects telemetry, applies behavioral analytics, and tracks metrics to identify anomalies and threats. Purple teaming unifies these perspectives through merged telemetry analysis, time-series scoring, and adaptive threat-hunting strategies. This stage ensures organizations maintain situational awareness, detect emerging risks, and continuously validate that monitoring systems are effective against real and simulated threats.

Red/Blue/Purple Teaming Requirements

- Synthetic user & rogue-agent generation
- Posture & metric collection
- UEBA for AI and agent signals
- Agent Behavior Analytics
- Purple SIEM lens
- Merged telemetry analysis
- Time-series scoring
- Adaptive hunt packs
- Model-drift vs threat-drift analysis

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

CHEAT SHEET genai.owasp.org

Red Teaming Landscape - Q2/3 2026

Govern

highflame
ADVERSA
ALICE
HIDDENLAYER
LASSO
TREND MICRO
Surelio.ai
humanbound
fiddler
TROJ.AI
NRI SECURE Deep AI Red Team
f5
tenuo
zscaler
opsin
VULCAN
NOMA
Straiker
FUJITSU
paloalto NETWORKS PRISMA AIRS
exabeam

Open Source Gold Sponsors Silver Sponsors

Governance ensures accountability, compliance, and continuous improvement. Red teaming supports auditability through reproducible attack-path evidence. Blue teaming aligns operations with regulatory and policy requirements while providing executive-level reporting on risk posture. Purple teaming enhances governance with residual risk analysis, forward-looking simulations, and integration of findings into retraining and incident response processes. This stage ensures that AI security is not just operationally effective but also measurable, auditable, and aligned with organizational and regulatory expectations over time.

Red/Blue/Purple Teaming Requirements

- Audit-grade attack-path replay
- Executive reporting
- Feedback to retraining & IR playbooks
- Framework mapping
- Policy & compliance orchestration (AI-TRiSM)
- Residual risk analysis cycles
- Risk simulators
- Signed artifact stores

- Refer to the **Red Teaming Solution Taxonomy Guide** for detailed descriptions
- Refer to the **Red Teaming Solution Capability Matrix** for specific solution capability coverage

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

https://genai.owasp.org/ai-security-solutions-landscape/

Gen AI Security Project Sponsors

Supporting Community Operations And Outreach Through Direct Financial Sponsorship

GenAI SECURITY PROJECT — GOLD PROJECT SPONSOR — TOP 10 for LLM and Generative AI

TREND MICRO · ByteDance · HIDDENLAYER · snyk · paloalto NETWORKS · Mend.io · Straiker

GenAI SECURITY PROJECT — SILVER PROJECT SPONSOR — TOP 10 for LLM and Generative AI

CHECK POINT · zscaler · Acuvity · f5 · SentinelOne · Cobalt

Starseer · securiti · Capsule · ActiveFence · NeuralTrust · TROJ.AI

GT cyber · PROTECTO · LASSO · CROWDSTRIKE · PROMPTARMOR · Synack · akto

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2

OWASP | GenAI SECURITY PROJECT

https://genai.owasp.org/ai-security-solutions-landscape/

Contributing to the Landscape Guide

For Red Teaming

Use the **QR Code** and associated form to submit a Red Teaming Security Landscape entry

Source; OWASP Gen AI Security Solutions Landscape Guide 2026. Q2