FLEXENTIAL

Penetration testing for AI systems

AI security starts with penetration testing.

Organizations are rapidly deploying AI systems—large and small language models, generative models, and agentic systems—which simultaneously introduce new security risks. These systems often operate with opaque logic, making them vulnerable to prompt injection, data poisoning, and model inversion attacks. Traditional security tools are ill-equipped to detect or mitigate these threats, leaving organizations exposed to adversarial exploitation, compliance violations, and reputational damage. As AI adoption accelerates, so do the risks:

- 72% of S&P 500 companies disclosed at least one material AI risk in 2025[1]
- 28% experienced attacks leveraging the AI application prompt[2]
- 84% of AI tools experienced data breaches; 51% faced credential theft[3]

Penetration Testing for AI Systems provides targeted, real-world security assessments tailored to AI systems. This service identifies vulnerabilities, delivers actionable remediation guidance, and strengthens cyber resiliency by empowering organizations to better secure their AI deployments and protect sensitive data.

How it works

Cyber resiliency through rigorous testing

Flexential's ethical hacker experts simulate real-world attacks on your AI systems to uncover vulnerabilities. We work closely with your IT and security teams to then independently assess threats, validate findings, and deliver comprehensive remediation guidance. Each engagement includes a detailed executive summary, prioritized risk analysis, and validation of remediation activities for improving your AI systems' security, compliance, and cyber resiliency.

> "Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks."
>
> **Stephen Hawking**
> Theoretical Physicist, Cosmologist, and Author

[1] https://corpgov.law.harvard.edu/2025/10/15/ai-risk-disclosures-in-the-sp-500-reputation-cybersecurity-and-regulation/

[2] https://www.gartner.com/en/newsroom/press-releases/2025-09-22-gartner-survey-reveals-generative-artificial-intelligence-attacks-are-on-the-rise

[3] https://cybernews.com/security/ai-tools-data-breaches-workplace-security-risks/

Problems we solve

Test your models to trust your models

- Model reliability
- Immature AI systems development
- Unidentified vulnerabilities in AI models
- Exposure to adversarial attacks
- Sensitive information leaking from AI systems
- Compliance gaps in AI governance
- Compromised AI systems generating poor outputs
- Compromised AI systems that allow for the stealing of corporate data

Key features

Expose (and remediate) vulnerabilities before attackers do

- Permission and authentication attacks
- Prompt injection and jailbreaking
- Supply chain compromise
- Adversarial input simulation
- Model fuzzing and inversion analysis
- Improper output handling
- Model theft testing
- Sensitive information disclosure
- Side-channel attacks
- Training data poisoning
- Insecure output handling evaluation
- Excessive agency or hallucination testing
- Identification of any denial-of-service conditions
- Remediation validation

Key benefits

Secure your AI systems

- Improved AI model security
- Reduced risk of data leakage and exfiltration
- Actionable insights for remediation
- Increased stakeholder trust in AI deployments
- Enhanced compliance posture
- Supports enterprise risk management

Outcomes we deliver

Protect your innovations

- Reduced organizational risk
- Improved AI output reliability
- Strengthened cyber resiliency
- Greater competitive advantage

> "AI models are susceptible to a range of threats—including adversarial attacks, data poisoning and model extraction—that can undermine their effectiveness or even turn them against their operators."
>
> "Organizations must adopt a zero-trust approach for AI systems, ensuring that every access is authenticated, model behavior is continuously validated and robust controls are applied to AI development and deployment environments. This scope should extend to all parties involved in AI system creation, from data sourcing platforms to fine-tuning services."

Abhijeet Mukkawar
Enterprise Architect, Siemens Digital Industry Software

Source: AI: Cybersecurity's Greatest Asset And Its Most Dangerous Threat