CASE STUDY

"At COE Security, we build, operate, and transfer world-class security solutions, empowering businesses with resilience and trust."

Joseph Henderson

COE SECURITY LLC

USA | BRAZIL | UAE | INDIA

CASE STUDY

SECURING AI & LLM SYSTEMS THROUGH PENETRATION TESTING

COE SECURITY

Measurable Impact Our Security in Numbers

- 1,500+ Global Engagements
- 95% Client Satisfaction
- 10+ Years of Cybersecurity Excellence
- 15K+ Critical Vulnerabilities Identified
- 1M+ Security Incidents Managed
- $350M+ Costs Saved
- $2.5B+ Transactions Secured
- 24/7/365 Threat Monitoring & Response
- Proactive Threat Hunting
- Cutting-Edge Security Technologies
- Dedicated Security Advisors
- Zero-Day Exploit Mitigation
- Incident Response Planning & Execution
- Security Awareness Training
- Customizable Security Solutions

Client Profile

A tech company specializing in AI-driven chatbots and enterprise automation relied on LLMs and AI APIs for business intelligence and customer support. To protect sensitive enterprise data, user interactions, and proprietary AI models, they needed to assess security risks in their AI/ML infrastructure, APIs, and compliance with GDPR, CCPA, and ISO 27001.

Challenges Faced

Before AI & LLM Penetration Testing, the company identified key security risks:

- Prompt Injection Attacks – Manipulating model responses to extract unintended data.
- Model Poisoning Risks – Injecting biased or harmful data into training datasets.
- Data Leakage Vulnerabilities – Exposing sensitive enterprise information through AI responses.
- Insecure API Endpoints – Allowing unauthorized access to AI services.
- Third-Party API Risks – Introducing supply chain security concerns.
- Model Output Manipulation – Subverting AI-driven decision-making.
- Adversarial Attacks – Forcing incorrect predictions with crafted inputs.
- Privacy & Compliance Gaps – Addressing GDPR, CCPA, and data retention issues.

Our Approach

To enhance the security of AI/LLM models, we conducted a comprehensive penetration testing engagement, identifying vulnerabilities and providing tailored remediation strategies.

We're Hiring!

COE SECURITY

Compliance & Regulatory Mastery

- ISO 45001
- HIPAA & HITRUST
- NIST & NIST 800-171
- ISO 27001
- PCI DSS
- CMMC
- CIS Controls
- SOC 2
- CCPA & NYDFS
- EU CRA
- FedRAMP
- GDPR
- UK Cyber Essentials
- Essential Eight - Australia

We're Hiring!

Scoping & Threat Modeling

We collaborated with the client to define:

- Scope of testing, including LLM-based applications, AI APIs, vector databases, training datasets, and model deployment environments.
- Threat models specific to AI/ML systems, such as data poisoning, prompt injection, API abuse, adversarial attacks, and data leakage risks.
- Testing methodologies, including Red Team AI Attacks, Black Box & Gray Box Testing, and Fuzzing Techniques..

Security Testing Execution

Using industry-standard frameworks like OWASP AI Security Guidelines, MITRE ATLAS, NIST AI RMF, and ISO 27001, we conducted AI/LLM security testing, covering:

- Prompt Injection Attacks – Evaluating AI resilience against manipulation.
- Data Exfiltration Risks – Identifying sensitive data leakage.
- Model Bias & Manipulation – Assessing input-based behavioral changes.
- Adversarial AI Testing – Using perturbation techniques to test robustness.
- AI API Security – Checking authentication, rate limiting, and access controls.
- Training Data Poisoning – Detecting vulnerabilities in model retraining.
- Model Drift & Hallucination – Analyzing unexpected AI deviations.

Findings & Risk Assessment

- After completing the penetration test, we provided a detailed AI/LLM security report, including:
- Critical, High, Medium, and Low-risk vulnerabilities, with business impact analysis.
- Proof-of-Concept (PoC) exploits, demonstrating how attackers could exploit AI models.
- A prioritized remediation roadmap, helping the company address AI security risks efficiently.

COE SECURITY

Our Cybersecurity Arsenal: Beyond Protection- We Build Resilience

- Application Penetration Testing (Thick/Thin)
- Mobile Application Penetration Testing
- API Penetration Testing
- Network Penetration Testing
- Operational Technology Security Testing
- Cloud Penetration Testing
- AI & LLM Security Audit and Pen Testing
- Red Teaming & Social Engineering Services
- Product & Hardware Penetration Testing
- IoT Security
- Security Operations Center Services (24/7)
- Custom Security Services

Remediation Support & Secure Hardware Development Best Practices

- Implementation of Guardrails & Content Filtering, reducing AI hallucinations and prompt-based exploits.
- Access Control Hardening for AI APIs, ensuring only authorized users interact with AI models.
- Input Validation Techniques, preventing adversarial and prompt injection attacks.
- Data Encryption & Masking, securing sensitive enterprise inputs.
- Monitoring & Logging Enhancements, improving AI model observability and security.
- Retraining Model Security Checks, ensuring malicious data is not used for fine-tuning.
- Implementation of Ethical AI Standards, improving AI transparency and compliance

Results

Within eight weeks, the company successfully:

- Eliminated all critical AI security vulnerabilities.
- Hardened LLM-based applications against prompt injections and API exploits.
- Secured AI-driven decision-making, preventing data manipulation risks.
- Implemented AI security best practices, ensuring ongoing AI model resilience.

We're Hiring!

COE SECURITY

Industries We Protect with Cutting-Edge Security



- Financial Services
- Healthcare
- Retail
- E-commerce
- Government
- Cryptocurrency
- Blockchain Technology
- Automotive
- Transportation/Logistics
- Energy & Utilities
- Hospitality/Tourism
- Entertainment/Media
- Manufacturing
- Education
- Telecommunications
- HiTech & Information Technology

We're Hiring!

Conclusion

By leveraging our AI & LLM Penetration Testing expertise, we helped the company identify vulnerabilities, enhance AI security, and ensure compliance with regulatory frameworks. Our structured approach, from threat modeling to remediation, ensured the AI-driven applications remained resilient against emerging cyber threats.

COE Security LLC

COE Security is a leading cybersecurity services provider, offering comprehensive solutions to address the evolving threat landscape. We have a proven track record of helping organizations of all sizes mitigate risks, strengthen defenses, and recover from cyberattacks. Our team of experienced cybersecurity professionals possesses deep expertise in the latest technologies and best practices, enabling us to deliver tailored solutions that meet your unique security needs.

"Peace of mind in a world of cyber threats. That's what we deliver."

Joseph Henderson

COE SECURITY