



COGNISYS

PENETRATION TESTING

LLM/Gen AI

Secure every interaction with our LLM/ Generative AI penetration testing.

In today's fast-evolving technological landscape, generative artificial intelligence (GenAI) and Large Language Models (LLMs) are actively reshaping industries, propelling innovation, and unlocking new avenues for commercial success. Nevertheless, these strides bring forth significant security challenges. Assuring the integrity and security of AI systems is imperative to upholding trust and defending sensitive data.

At Cognisys, we recognise the unique security needs of AI providers, producers, and customers. Our LLM Penetration Testing service is meticulously designed to address these requirements. They provide specialised security assessments tailored to each role within the AI ecosystem, ensuring the safety and integrity of your AI systems and the data they process.

Why get an LLM penetration test?

Progressively integrated into diverse applications, AI applications driven by Generative Language Models (LLMs) stand as susceptible targets for cyber threats. Our penetration testing services are meticulously tailored to safeguard your AI models, fortify the data they handle, and uphold their integrity.

We adhere to a comprehensive strategy founded on the LLM-OWASP Top 10 framework, an established methodology adept in identifying and mitigating vulnerabilities, therefore furnishing you with a stalwart defence mechanism against potential threats.

Our approach

Our team of experienced ethical hackers, armed with advanced techniques, replicate real-world attack scenarios on your LLM/GenAI systems. We follow a black box testing approach, engaging with the AI as regular users to discover concealed vulnerabilities. This frequently involves interacting with chat or voice applications to test the system's susceptibility to manipulation, data extraction, and model integrity compromise.

Each penetration test we conduct includes the following:

- **Live hacking events:** Our team performs live hacking sessions to demonstrate real-time vulnerabilities and their potential impacts during our project debriefing.
- **Detailed reports:** Our thorough reports highlight critical weaknesses and provide actionable recommendations for your IT team to implement, empowering you to strengthen your security.
- **Mitigation strategies:** Beyond identifying vulnerabilities, we offer strategic guidance to help you swiftly mitigate risks and improve your security posture.

Unmatched precision and quality

We have developed a thorough framework for LLM penetration testing to stay ahead of evolving threats and ensure the strength and resilience of your AI systems against attacks.

The importance of proactive security

As GenAI continues to advance, so do adversaries' tactics. The accelerated development of AI technologies necessitates proactive security measures to stay ahead of potential threats. By partnering with us for your LLM penetration testing needs, you can access industry-leading expertise and cutting-edge techniques that fortify your AI deployments, ensuring you are aware and prepared for the evolving threat landscape.

Our LLM/Generative AI penetration testing methodology

Cognisys has pioneered a highly detailed LLM Penetration Testing methodology, uniquely tailored to meet the most current standards and best practices in AI security, such as the OWASP Top 10 for LLM applications.

[BACK TO CONTENTS](#)

