Press release

Add value.
Inspire trust.

Penetration tests for AI systems

December 3, 2025

TÜV SÜD tests IT-Security of Artificial Intelligence

Munich. TÜV SÜD supports companies by conducting penetration tests for AI systems. Maintaining network security by combating phishing, ransomware attacks, and DNS hijacking has long been part of everyday life for CISOs, as attacks via AI systems are becoming increasingly common. Cybercriminals use methods such as prompt injection and model inversion. They also target companies' sensitive data and trade secrets. AI penetration tests ensure the integrity, fairness, and robustness of AI systems.

AI penetration tests are specialized assessments designed to uncover gaps in AI and machine learning systems. They evaluate the robustness of the model, the security of the data pipeline, and the vulnerability to threats such as adversarial attacks, model inversion, or data poisoning. The goal is to ensure the confidentiality, integrity, and reliability of AI-powered applications under realistic attack scenarios.

Application scenarios

There are various application scenarios for AI penetration testing: They can help identify LLM security vulnerabilities in web applications early in the development phase, assess the risk of data leaks, misuse, or manipulation of LLMs in applications, or comprehensively review predictive and user-defined LLMs in terms of data, training, and algorithms.

Five steps to more robust AI systems

Experienced AI experts from TÜV SÜD support companies in improving their AI systems in five steps. In a kick-off meeting, goals are defined and processes discussed. Then, relevant information is gathered. The third step is the actual pentesting. Unlike traditional pentesting of networks or servers, AI pentesting requires knowledge in the areas of machine learning, testing input/output behavior, and model logic. TÜV SÜD bases its assessments on the NIST AI Risk Management Framework, OWASP Top 10 for LLMs/ML Security, and MITRE ATLAS testing standards. The results are then analyzed by TÜV SÜD and finally discussed with the company.

“The more widespread AI becomes, the more attractive it becomes to criminals. Due to its rapid development, security is often not yet optimally integrated. Whether companies are optimizing AI models or integrating LLMs into applications, AI system-specific vulnerabilities must be identified at an early stage,” says Vaibhav Pulekar, Senior General Manager Cybersecurity at TÜV SÜD. "Those who do not have their models or applications checked risk serious security and data protection risks due to gaps in the system."

Further information on TÜV SÜD's AI services is available at tuvsud.com/en/topics/artificial-intelligence.

Media Relations

TÜV SÜD Public Relations Westendstraße 199 80686 Munich	Laura Albrecht Phone +49 89 5791-2935 Email laura.albrecht@tuvsud.com Website tuvsud.com/newsroom

Founded in 1866 as a steam boiler inspection association, TÜV SÜD is now a global company. Around 30,000 employees at over 1,000 locations in around 50 countries ensure the optimization of technology, systems, and expertise. They make a significant contribution to making technical innovations such as Industry 4.0, autonomous driving, and renewable energies safe and reliable. tuvsud.com/en