AI Security Testing Services

As with any rapid technological advancement, new challenges and threats arise – many of which have already surfaced with AI and Machine Learning (ML).

Securing AI and ensuring the safety and security of people, processes, and technology in an AI-augmented world demands vigilance and a commitment to forward-thinking strategies that require continuous adaptation.

Managing AI Cyber Threats

With comprehensive AI/ML security assessments, NCC Group combines AI/ML expertise with industry-leading penetration testing enabling organisations to make informed decisions on how best to safeguard valuable data and secure intelligence from the dynamic landscape of AI cyber threats.

Our range of services focus upon:

Benchmarking resilience against attacks, the likelihood and impact of vulnerabilities in line with policy compliance and AI security best practice, with remediation recommendations.

Evaluating AI/ML models for safety and alignment, considering trends of bias, to measure, analyse and remediate the risk of hidden misalignment to the model's intended purpose.

Comprehensive assessment of AI/ML integrated capabilities and cloud infrastructures, to identify system security weaknesses and their consequences for organisations' people, processes, and technology.

The potential compromise of model-governed assets for example: training data, model weights, prompt secrets, deepfakes, API endpoints, plugin endpoints, and more.

Improving the state of security: our experts provide both technical and strategic recommendations for both short-and-long term improvements.

Assessments aligned with the MITRE ATLAS framework and OWASPs Top 10 AI/ML Vulnerabilities: including, but not limited to: Prompt Injection, Adversarial Attacks, Data Poisoning Attacks, Membership Inference, Model Inversion and Model Stealing.

Our AI Services

Red Teaming
Our AI/ML experts identify security weaknesses with a tailored approach that targets your environment. We will test against known best practices to expose AI/ML specific vulnerabilities such as OWASP's AI/ML and LLM Top Ten.
Our team will work with you to explore any potential threats introduced by the use of the AI/ML systems. Post-assessment, we provide a detailed report including recommendations, to help you bring your environment in line with security best practice.

Cloud Security Review
Evaluate the configurations of your AI/ML cloud or third-party infrastructure. Our experts will examine key configuration parameters of the specific AI/ML cloud infrastructure that impacts upon system security. Our reviews provide technical findings and strategic recommendations to improve the state of your security in both the short and long term.

Threat Modelling
Our consultants collaborate with your organisation's subject matter experts (SMEs) to review design and architecture decisions, assess risk profiles, and evaluate the overall security posture of your AI/ML-integrated environment. During this process, we conduct thorough implementer interviews and meticulously evaluate various aspects of your platform, including assets, controls, data flows, trust boundaries, and threat actors.

Secure Development Lifecycle Testing
Secure solutions come from robust policies and procedures. Our experts will work with you to understand and analyse your current Secure Development Lifecycle (SDL) AI/ML processes, policies and production pipeline. We will work with you to identify and address the security holes in your AI/ML SDL.

Bias and Toxicity Assessment
ML models are trained on flawed data, and often hide unpredictable qualities that introduce risks. We measure and analyse generative and non-generative AI models for safety and alignment, to remediate the risk of hidden misalignments within models. Our experts evaluate behavioural trends to identify dangerous patterns potentially indicating biases embedded into performance, and provide recommendations to isolate and eliminate adverse outputs for Responsible AI practices.

Deepfake Vishing
Deepfake Vishing leverages AI-generated deepfake voice technology to impersonate real individuals to scam victims. This sophisticated technique involves attackers using (ML) models trained on real voice samples to create highly realistic synthetic voices that mimic tone, accent, and speech patterns. Our service includes creating voice models to clone specific members of staff voices used to simulate threats targeting your organisation and policies to detect and defend against such attacks.

Why NCC Group

NCC Group is passionate about sharing our insights and intelligence from operating at the 'frontline' of cyber security with policy makers who are making important decisions about the future of AI. We have engaged with governments, regulators, and legislators across the world, helping to inform new laws and regulations and advocating for a more secure digital future. Recent highlights include inputting into the Australian Federal Government's discussion paper on AI regulation, providing evidence to a UK Parliament inquiry on LLMs and supporting the development of the UK's AI whitepaper.

Track Record
Rich heritage of AI/ML security testing, including AI specific components within M&A over the last 10 years.

Experience
Achieve optimum enterprise cyber security resilience with NCC Group's highly skilled AI/ML security experts.

Research-led
World class AI/ML security research, fueling NCC Group's expert delivery methodology.

Technical Assurance
Collaborate with NCC Group's experts and assess AI/ML-driven product solutions and operations in the wider context of AI/ML business utilisation.

Actionable Intelligence
Guide decision-making with clear, actionable recommendations for secure intelligence with evolving technologies.

Manage Compliance
Aligned with regulation and compliance criteria for AI/ML technologies, and processes.

For more information from NCC Group on our G-Cloud 15 service, please contact:

+44 (0)161 209 5200

bidteam@nccgroup.com

© 2026 NCC Group. All rights reserved. Please see www.nccgroupplc.com for further details. No reproduction is permitted in whole or part without written permission of NCC Group. Disclaimer: This content is for general purposes only and should not be used as a substitute for consultation with professional advisors.