CREST study: AI is now routine in threat-led penetration testing



CREST research finds 81% of surveyed providers use AI in threat-led penetration testing. Autonomous testing is still rare, and clients should expect more AI disclosure and approval requirements.

- 81% of 26 surveyed providers use AI in TLPT beyond experimentation; 65% expect it to support 40% or more of tasks within 12 months.
- None of the 11 regulated TLPT providers that answered allow AI to execute tests autonomously.
- 92% say sending client information to external or cloud-hosted AI services needs explicit client approval.
- 85% expect more AI-related procurement requirements and 81% expect more requests to disclose AI use.

CREST has published research on how AI is changing threat-led penetration testing (TLPT). It draws on responses from 26 providers and two industry roundtables, with a focus on financial services.

What the research found

AI is already part of delivery: 81% of surveyed providers use it beyond experimentation. Reporting and assessment is the most common live use, followed by reconnaissance and threat-informed preparation. Of providers using AI to emulate threat actors, 86% report better accuracy and realism.

The report also flags risks. Model and training-data bias can produce generic patterns that blur differences between adversaries. AI-enabled impersonation, such as voice cloning or synthetic video, needs justification, authorization and safeguards proportionate to potential harm.

Control and data protection

Autonomy is limited. None of the 11 regulated TLPT providers that answered the question allow AI to carry out testing activities autonomously. Scope, operational control, and client data and confidentiality were each ranked a top-three delivery risk by 68% of respondents.

On data, 92% say submitting client information to external or cloud-hosted AI services requires explicit client approval. Relevant considerations include processing and storage locations, retention, model training practices, subcontractors and access rights. The report also notes that human approval steps can slow simulated attacks and affect conclusions about detection and response, so changes to test conditions should be documented.

What to expect

Providers expect more AI-generated offensive tooling, and 85% expect stricter AI-related procurement requirements. Institutions commissioning tests should be ready to ask clear questions about how AI is used and governed in their engagements.

What to do now

1. Ask TLPT providers to disclose where and how AI is used in your engagements, including any external or cloud-hosted services.
2. Require explicit written approval before any client data is sent to external AI services, and define storage, retention, model-training and subcontractor terms.
3. Set rules of engagement that specify whether autonomous AI actions are permitted, with stop conditions and logging of automated actions and human interventions.
4. Require authorization and safeguards for AI-enabled impersonation in social-engineering exercises.
5. Add AI-related clauses to penetration testing procurement and review how approval delays affect test conclusions.