AI and LLM penetration test scoping questions

These are the questions a SilentGrid consultant asks when scoping an AI or LLM penetration test. Answering them first makes the call shorter and the quote more accurate.

"Not sure" is a fine answer. The scoping call helps resolve unknowns.

Questions marked **Sizing** help us prepare an accurate quote.

01 Before any test

What prompted the test, and what date does the result need to meet?

Why we ask: the reason sets the depth and report format, and the date fixes the testing window and any retest.

Will testing run against production, staging or a dedicated test copy?

Why we ask: production needs tighter limits on load, data and timing. A test copy allows more thorough techniques.

When can testing run, and are there change freezes or peak periods to avoid?

Why we ask: the schedule is built around them.

Will you want fixes retested?

Why we ask: retesting is agreed at scoping, so it is scheduled and priced from the start.

02 AI and LLM systems

What kind of system is it: a chatbot, an assistant over internal data, an agent that takes actions, or a model you trained?

SIZING

Why we ask: each has a different attack surface and test plan.

What can it reach: documents through retrieval, tools, APIs, customer records, email?

Why we ask: prompt injection is only as serious as the access behind it.

Can it act on a user's behalf: send messages, change records, run code?

Why we ask: tool access turns a successful injection into an unauthorised action.

Does it read content from untrusted sources: web pages, uploaded files, inbound email?

Why we ask: indirect prompt injection arrives through content the model reads.

Which model and hosting: a provider API such as Azure OpenAI or Amazon Bedrock, or self-hosted? Has it been fine-tuned?

Why we ask: self-hosted and fine-tuned models add model theft and data poisoning to the scope.

What must it never do or disclose?

Why we ask: guardrails are tested against those specific rules.

Are there rate limits or per-call costs?

SIZING

Why we ask: testing sends a large number of prompts, so limits and costs need agreeing first.

Discuss your testing priorities

www.silentgrid.com/contact