

Build vs. Buy: AI Penetration Testing

What it really takes to build, run and maintain AI pentesting in-house.



Building In-House: Easy to start. Hard to trust and manage at scale.

Frontier models make it easier to find vulnerabilities. The hard part is turning them into AI pentesting you can trust. You have to validate every finding, test safely, control costs and keep up with changing models That takes a team, a budget, and months of work to build and maintain, while attackers are already using AI to exploit faster.

Meet Novee: AI pentesting, ready to run

Novee is an AI penetration testing platform delivering deeper findings, higher precision, built-in safety, and predictable costs at enterprise scale. With Novee, you can start testing now, without the time, cost and risk of building and maintaining a platform on your own.

The Real Build vs. Buy Question

You can build an AI pentesting agent. The question is whether you want to take on the ongoing cost, risk, and engineering required to operate a continuous offensive AI testing platform.

	Buy – Novee	Build In-House
COST		
AI costs	✔ Predictable pricing, with cost continuously optimized	⚠ Token spend grows with every run, app and test
Infrastructure	✔ Containers, proxy, storage and connectors run for you	⚠ Build, run and secure the full stack yourself
Staffing	✔ Dedicated TAM with hands-on support	⚠ Scarce offensive security, ML and platform engineers, pulled from core work
Maintenance	✔ Models and orchestration continuously maintained	⚠ Constant re-tuning, and the effort stalls if key engineers leave
RISK		
Validation	✔ Validated findings with proof of impact and PoC scripts	⚠ Build validation, reproducible proof, and PoC scripts yourself
Safety	✔ Layered guardrails, tested against a safety benchmark	⚠ Build and prove your own proxy, intent checks and safety benchmark
Visibility	✔ Full agent trace, traffic and audit log	⚠ Design and instrument all of it yourself
TIME & CAPABILITY		
Speed to value	✔ Live today. Assessments in 6 to 48 hours	⚠ Quarters to build, longer to trust in production
AI performance	✔ Best model for each task, plus a proprietary model	⚠ Pay for multiple models and keep benchmarking them yourself
Application intelligence	✔ Testing gets deeper and more efficient with every assessment	⚠ Build a way to retain and reuse what every test learns
Integrations	✔ Built into CI/CD, ticketing, and security workflows	⚠ Build and maintain every integration yourself

Why It's Hard to Build Yourself

Three problems every in-house build runs into, and what it takes to solve each one.

01 Proprietary Model + Multi-Model Routing

Models Change Every Month

⚠ To build it yourself

Train and own a model, re-benchmark every new frontier model as it's released, retune your prompts, and route each task to the model that performs best.

✅ What Novee built

Novee post-trains its own offensive model and combines it with frontier models. Every task goes to the model that performs best, with no rework on your side.

02 Novee Gym

Continuously Proving It Works

⚠ To build it yourself

Generate test environments with known vulnerabilities, a scoring engine and a loop that feeds results back into the system. Without them, you can't tell if a change helped or hurt.

✅ What Novee built

The Novee Gym tests the full Novee offensive AI system against thousands of real applications with known vulnerabilities. It then optimizes it for precision, recall, quality, safety, speed and cost.

03 Asset Intelligence Model

Understanding & Remembering How Each App Works

⚠ To build it yourself

Create a way to store what each test learns and reuse it, so testing gets deeper instead of starting over.

✅ What Novee built

Novee's Asset Intelligence Model keeps a persistent understanding of each app's workflows, permissions, APIs and business logic, so testing gets deeper, more targeted and more efficient with each assessment.

5 Questions to Ask Before You Build

- 01 How will you independently prove findings are real and measure false negatives?
- 02 How will you prevent an autonomous offensive agent from taking unsafe actions?
- 03 How will you measure and control model, token, compute, and infrastructure costs at scale?
- 04 Who will own, operate, and continuously improve the system after it launches?
- 05 How long until it's production-ready – and what protects you until then?

NOVEE IS CHOSEN BY TEAMS THAT TAKE ATTACKERS SERIOUSLY



Book a demo

Start Testing With Novee Today

Run Novee against your applications and benchmark the results before committing a team to building in-house.