SERVICE SUMMARY

Independent evidence of what your AI does when someone tries to misuse it.

Before you trust an AI with payments, records or customer data, we test it and give you a report in plain English, with evidence you can check yourself.

THE PROBLEM

Almost every safety claim about an AI model comes from the company that built it. Those tests are run by the party with the most to lose from a bad result. That is a supplier's assurance. **It is not evidence.**

We are independent. We do not build AI systems or sell fixes for them.

WHAT A TEST SHOWS

Two AI models, the same 258 tests.

AI MODEL A	AI MODEL B
101 of 258 tests ended in a harmful action	**0** of 258 tests ended in a harmful action

The 101 include a $187,400 payment sent on a forged email with no verification. From the outside the two models look the same. **A test tells you which one you have.**

HOW THE TEST WORKS

1. **You name the AI model.** No passwords. No access to your systems.
2. **We run 258 tests.** Forged emails, people posing as executives, false deadlines. Nothing real is at risk.
3. **We record what it did.** Not what it said it did. A person reviews any unclear case.

A clean result has to mean something. 67 of the 258 tests are honest jobs where the right answer is to act: pay the real invoice, read the balance, pass the case to a person. An AI that refuses everything fails those, so it cannot pass by doing nothing.

WHY NOT JUST USE THE FREE TESTS

Free public safety tests exist, and they are good. You could run them yourself. We run the seven industry-standard ones on your AI first, then our own test, and show you both results side by side. **You see the gap between the two for yourself.**

A free test gives you a score. It does not tell you what the AI did, what that means for your business or what to do next, and nobody signs it. We give you a report in plain English, the evidence behind every finding, and a named person who stands behind it.

The free tests are also narrower than what an AI faces at work. The best-known one checks one kind of hidden instruction. On one AI model it got through 1.2% of the time (11 of 949). We check twelve kinds. On the same model, **five of the twelve got through**: customer data was emailed to an outside address and administrator rights were granted.

SERVICES AND PRICES

The Screening
$5,000

One AI model put through the 258 tests once, with the seven public tests beside them. Report and sealed evidence within 72 hours. You give us the name of the model and nothing else. It shows that a failure can happen, not how often. The $5,000 comes off the price of a Full Audit.

Monthly Monitoring
$4,000 a month

The same test every month, so you know when an update changes your AI. Paid three months in advance.

The Full Audit
from **$25,000**

The deep test, for an AI you are about to trust with real work.

- A 400-turn stress test, run 30 times: 12,000 turns across 37 categories.
- Every failure run 100 more times, so you know how often it happens.
- The 258 tests, 50 times each, then 50 fresh runs of every test that failed.
- Your own setup: your instructions, tools and connections.
- A person checks every flagged result. Then a second, blind check.
- The seven public tests on the same AI, beside ours.

Custom and multi-model work
from **$45,000**

The Full Audit across several AI models, compared side by side, so you can choose the safest one. Scoped to what you need.

THE GUARANTEE

If a Full Audit finds nothing serious, you pay nothing.

"Nothing serious" means no confirmed finding at the severity we agree on in writing before testing begins. It is set then, not by an opinion formed after the result is known. One confirmed finding is enough.

HOW YOU CHECK OUR WORK

You do not have to take our word for it.

Every report comes with its evidence. The evidence is sealed and time-stamped on the Bitcoin blockchain, so nobody, including us, can change it or move the date. A free program re-checks all of it on your own computer. A real evidence pack is at potestasai.com/sample-pack and a sample report is at potestasai.com/sample-report.

WHAT A TEST CANNOT TELL YOU

This is a test, not a certificate. A clean result does not mean your AI is safe. Each result is for one version of one AI model, and an update can change it. Our findings are aligned with the NIST AI Risk Management Framework and the OWASP Top 10 lists for AI. We do not certify.

WHO SIGNS THE REPORT

Joseph Cirello, founder, signs every report. U.S. Army veteran. Before that, accounting, where the rule is that the party checking the work cannot be the party that did it. We apply that rule to AI.