TVARIS LABS · HYDERABAD, INDIA

Reliability is the new bottleneck in AI.

The model is no longer the hard part. Keeping an agentic system correct, affordable and debuggable on its ninetieth day in production is. That layer is what we build.

SYSTEMS OPERATED	WHO YOU WORK WITH	CLIENT ENGAGEMENTS
Five	Both founders	You would be first

01 / The shift

The hard part moved, and most roadmaps did not move with it.

For three years the binding constraint on an AI feature was model capability. You waited for a better model and your problem got easier. That era closed. Frontier models now clear the bar for the overwhelming majority of commercial work on the first try.

What did not get easier is everything around the model. A system that calls tools, holds state across a long task and takes real actions has a failure surface that looks nothing like a request-response API. It degrades quietly. It costs a different amount every day. A prompt edit on Tuesday can break an output path nobody looks at until Friday.

So the question stopped being *can the model do this* and became *can we tell, on any given day, whether it still is*. That is an engineering problem, not a research one, and it is answered with the same boring apparatus every other production system needs: measurement, telemetry, containment, and data you can trust upstream of all three.

---

02 / What the shift costs

Two teams ship the same feature. A quarter later they are not in the same place.

The difference is rarely the model, the framework or the talent. It is whether anyone built the instrumentation before the system got complicated enough to need it.

WITHOUT THE LAYER	WITH IT
Prompt changes ship on intuition; nobody can prove the system got better.	Every change is scored against a regression set before it reaches anyone.
Quality regressions are discovered by users, weeks after the commit that caused them.	Quality drift shows up on a dashboard, not in a support ticket.
Spend is a monthly surprise with no per-feature attribution behind it.	Cost is attributable to a feature, a tenant and a code path.
One flaky upstream call takes down the whole task instead of one step.	Failures are contained to the step that failed and retried on a known policy.
The team slows down deliberately, because every change is a gamble.	The team ships faster over time, because changes stop being gambles.

The compounding runs in both directions. Instrumentation added in month one is a week of work. Added in month nine, after the system has grown branches nobody remembers, it is a rewrite with a live product on top of it.

03 / What you get to have

A system you can change on a Tuesday.

> **An AI system where any engineer can make a change, see within minutes whether it made things better or worse, and know what it will cost.**

That is the state we are trying to put you in. Not a smarter model — a system whose behaviour is legible enough that improving it is ordinary engineering work rather than an act of faith.

It is difficult to reach alone, and the reason is structural rather than a matter of skill. The work is invisible until it is missing, it competes directly with feature deadlines, and it demands a combination — data engineering at scale, evaluation methodology, production operations — that rarely sits inside one person or one team.

04 / What we are

An engineering studio, working on the half nobody demos.

For **product and data teams putting an agentic feature into production**, Tvaris Labs is an **engineering studio** that builds **the evaluation, telemetry and data systems that keep it correct once it is live.** Unlike a consultancy staffing engineers by the month, or a platform selling you another dashboard to configure, **we deliver working systems we have already built and operated for ourselves.**

The realistic alternatives are a generalist agency that has not run one of these in production, an observability vendor whose product starts where your problem ends, or your own team doing it between sprints. We are useful specifically when the third option keeps losing to the roadmap.

05 / How we get you there

Four capabilities. Each one removes a specific way the Tuesday change goes wrong.

01

Evaluation harnesses

Regression sets, graded rubrics and LLM-as-a-judge scoring wired into the deploy path, so a change is scored before it reaches a user. **Removes:** shipping on intuition.

02

Cost and quality telemetry

Token, latency and outcome traces attributed to feature, tenant and code path, with drift alerting on the metrics that matter. **Removes:** the monthly billing surprise and the silent regression.

03

Agent orchestration and containment

Tool use, long-running state, retry and fallback policy, and approval gates wherever a wrong action would be expensive to undo. **Removes:** one failed step taking down the whole task.

04

The data underneath

Ingestion, transformation and cost control on Azure and AWS — our deepest bench, built at petabyte scale in production before this company existed. **Removes:** a well-instrumented system reasoning over bad input.

06 / Why believe any of this

We have no client logos. Here is what we have instead.

Stating this plainly is deliberate. You would be our first client engagement, and that belongs near the top of the page rather than in a late discovery on a call.

Five systems we built and still operate

DataCraves, RoboLLMs, Tvaris, FocusTu and ShortML are live, with real authentication, billing, entitlements, deployment and evaluation behind them. Not prototypes. We run them, including the parts that page us.

Decisions we can walk you through, including the wrong ones

For any of the five we can show you the architecture, what it cost, what we would build differently now, and the incidents that changed our mind. That is a more useful signal than a testimonial and harder to fake.

Depth in the specific problem, before the studio existed

Eight years of production data and ML platform work at Microsoft, Amazon and JPMorgan Chase — including pipelines over 200 TB of daily logs, a 70 percent compute and storage reduction, and an LLM-as-a-judge evaluation harness for a shipped Copilot feature.

Prices and limits published before you ask

Product pricing is on the site. What we will not take on is written down on the home page. Both are claims you can verify in one second, which is more than a reference call gives you.

What we are not claiming: users we do not have, revenue we have not earned, or clients we have not served. If a number is not on this page, it is because we could not stand behind it today.

07 / The ask

One conversation, thirty minutes, no invoice attached.

You describe the system and where it is fragile. We tell you what we would instrument first, and whether the problem is actually one we should touch. If it is a data problem or a process problem wearing an AI costume, you will hear that on the call rather than in month six.

Direct, no form funnel: contact@justaskstuff.com

Tvaris Labs LLP LLPIN ACZ-2096 GSTIN 36ABAFT3121C1ZS Rajapushpa Atria, Kokapet, Hyderabad, Telangana, India

Reliability is the new bottleneck. That layer is what we build.

contact@justaskstuff.com

justaskstuff.com

Hyderabad, India

Tvaris Labs LLP · LLPIN ACZ-2096 · GSTIN 36ABAFT3121C1ZS
Rajapushpa Atria, Kokapet, Hyderabad, Telangana, India
Current pricing and availability at justaskstuff.com/pricing