

LAB L1

BETA

MULTI-AGENT SYSTEMS

# Agentic Workflow Studio

Planner - Critic - Tool Registry - Eval Harness

A composable framework for building, evaluating and governing multi-agent systems in production.

Thesis: Agents need the same engineering posture as microservices: contracts, evals, traces, blast-radius controls and rollbacks.

TASK COMPLETION

**+ 4.2**

HALLUCINATION RATE

**81%**

COST PER RESOLVED TASK

**64%**

## MANIFESTO

### Why this lab exists

Most agent demos collapse in production because they ship a planner without a critic, tools without a registry, and evals without a regression suite. The Studio treats agents like software: typed contracts, deterministic fallbacks, golden sets, signed tool calls, and a tracing plane your SRE team actually trusts.

## KPIS

### Outcomes we measure

- Task completion: + 4.2
- Hallucination rate: 81%
- Cost per resolved task: 64%

## ACTIVE EXPERIMENTS

### What the lab is testing now

#### > Planner/critic loops at scale

Comparing reflexion, tree-of-thought and graph planners on 12 enterprise task suites.

#### > Deterministic fallbacks

When confidence dips below threshold, route to typed deterministic code paths - measured cost & accuracy uplift.

#### > Cross-model arbitrage

Real-time routing between GPT-5, Claude, Gemini 3 and OSS Mistral based on task class, cost ceiling and region.

#### > Agent blast-radius limits

Per-tool spend and side-effect quotas enforced by the runtime, not by prompts.

## SHIPPABLE ARTEFACTS

# Everything that ships

---

### > Studio SDK

TypeScript + Python SDK for declaring agents, tools, evals and traces in code.

### > Eval harness

Golden sets, regression suites, drift detection, CI gates that fail builds on quality dips.

### > Tool registry

Signed, versioned tools with auth scopes, cost ceilings and per-tenant policy.

### > Tracing plane

OpenTelemetry-native traces of plan, tool calls, model routes and cost - exportable to Datadog / Honeycomb.

### > Reference patterns

Six production blueprints: renewals, support, research, ops, claims, recruiting.

## LAB TEAM

# Who you'll work with

---

- Agent Systems Principal
- Eval Engineer
- Tool / SDK Engineer
- Tracing & Observability Lead

## ENGAGEMENT TIMELINE

# Weeks 1-8 - first agent in shadow by week 2

---

### 1 Weeks 1-2 - Shadow deploy

Wrap an existing workflow; capture baseline cost, latency and quality with the eval harness.

### 2 Weeks 3-5 - Planner + critic + tools

Wire planner/critic loop, register typed tools, ship deterministic fallbacks.

### 3 Weeks 5-8 - Production hand-off

Tracing, blast-radius limits, model routing, runbook + SRE on-call rotation.

## FLAGSHIP PODS

# Squads that productionise this lab

---

- Renewals Agent Pod
- Support Triage Pod
- Research Co-pilot Pod
- Claims & Adjudication Pod

## PARTNERS

# Who we collaborate with

---

OpenAI - Anthropic - Google DeepMind - LangChain - DSPy - Datadog - Honeycomb

PUBLICATIONS

## Receipts

---

### Critic-grounded planners outperform monolithic agents on enterprise tasks

NeurIPS Workshop on Agentic AI - 2025

### Deterministic fallbacks: a cost-quality study across 1.4M agent runs

AXP Internal Whitepaper - 2026

FAQS

## What partners actually ask

---

### Q. Is this another agent framework?

A. No - it's the missing engineering layer around frameworks. Use LangGraph, DSPy, OpenAI Agents or your own; the Studio adds evals, tools, traces and policy.

### Q. How do you stop agent loops blowing budgets?

A. Per-tool and per-task spend ceilings enforced at the runtime, plus auto-cancellation at policy thresholds. Logged to your FinOps cockpit.

### Q. Can it run on-prem?

A. Yes - the runtime ships as a container; OSS models served on your private GPUs are first-class.

### Q. Do you replace our SRE / observability stack?

A. No. Traces export via OpenTelemetry to whatever you already use (Datadog, Honeycomb, Grafana).

## Partner with Agentic Workflow Studio

Outcome-priced. Sovereign by default. Refund-backed if the contracted KPI isn't hit.

Apply: [alfaxprienz.com/labs/agentic-workflow-studio#partner](https://alfaxprienz.com/labs/agentic-workflow-studio#partner)