Jetty

AI/ML workflows that check their own work

Build · Run · Monitor — from any AI coding agent

AI agents fail silently — and no one knows until it's in production

The enterprise AI stack has a missing layer

The stack exists — the glue doesn't:

- **Models** are deployed: GPT-4o, Claude, Gemini
- **Agent frameworks** are proliferating: Claude Code, Cursor, LangChain
- **Cloud infra** is ready: AWS, GCP, Azure

What's missing: a layer that evaluates, orchestrates, and systematically improves AI workflows before they reach users

Evaluation isn’t a feature — it’s the foundation

Jetty: the agentic evaluation platform for AI/ML workflows

One platform for the full workflow lifecycle:

- **Build** — define workflows as versioned JSON configs with typed inputs and modular steps
- **Run** — execute synchronously or async from any MCP-compatible agent
- **Monitor** — every step recorded as a trajectory with artifacts and annotations
- **Improve** — iterative refinement with built-in quality gates

Two feedback loops, one platform

Inner loop

Each run self-corrects:

- Quality gates evaluate each step
- Up to 3 refinement iterations per run
- Structured evaluation criteria baked in
- Run fails fast or improves to threshold

Outer loop

The system improves across runs:

- Every trajectory labeled and indexed
- Compare runs across models and configs
- Patterns surface automatically
- Teams build institutional knowledge

MCP-native: works with every agent, from day one

16 MCP tools for workflow management — no custom integrations required:

- **Claude Code · Cursor · VS Code Copilot**
- **Gemini CLI · Windsurf · Zed · Codex CLI**
- Any MCP-compatible agent (the standard is winning)

MCP is becoming the universal agent interface. Jetty was built for this moment.

Runbooks: structured execution with built-in quality gates

Runbooks are parameterized markdown documents that tell an agent *how* to complete a complex task:

- Step-by-step execution with explicit quality checkpoints
- Up to 3 refinement iterations per gate
- Parameterized inputs — run the same workflow with different data
- Secrets resolved through a secure 3-tier fallback system

"A structured markdown document that tells a coding agent how to accomplish a complex, multi-step task with built-in evaluation loops." — jetty.io

Full trajectory: every step recorded, labeled, downloadable

Every workflow run produces a complete trajectory:

- Step-by-step outputs captured in sequence
- Artifacts attached and downloadable
- Annotations and quality labels per run
- Ask any question about any prior run, instantly

No more ”what did the agent actually do?” — full observability at every depth.

Already deployed at AWS, Google, and frontier AI labs

Customers spanning research and enterprise:

- **AWS · Google · TU/e** (Eindhoven)
- **Brickroad AI · Akinox · Carepath**
- **OpenML · AI Vibe · Workshop.ai**

Use cases: operations automation, competitor research, engineering audits, document processing

5-tier SaaS · Free → Enterprise

Free — 100 runs/mo, self-serve adoption

Builder — $200/yr, 10,000 runs/mo

Team — $800/yr, 350,000 runs/mo, 15 seats

Business — $4,800/yr, 1M runs/mo, 50 seats

Enterprise — Custom · SOC 2 · HIPAA · on-premise

The market: every enterprise deploying AI agents needs this layer

The evaluation and orchestration gap is universal:

- Every team running LLM workflows faces the same reliability problem
- MLOps, AI observability, and agent orchestration are merging into one category
- The window to define that category is open — and closing

Jetty sits at the intersection and is building the standard before it's set.

Go-to-market: agent ecosystem as the distribution channel

Wedge: MCP integration puts Jetty in every agent's toolbox at install

Land: Self-serve free tier — developers adopt, workflows proliferate

Expand: Teams, departments, then enterprise contracts with custom SLAs

Moat: Trajectory data accumulates; switching cost grows with every labeled run

Backed by Hidden Layers, AQC Capital, and Mila Ventures

Strategic alignment across AI infrastructure and enterprise software:

- **Hidden Layers** — AI infrastructure specialist fund
- **Mila Ventures** — connected to world-class AI research (Yoshua Bengio's lab)
- **AQC Capital** — enterprise software depth and network

Why now: the MCP moment is here

The agent protocol is standardizing — fast:

- Claude Code, Cursor, VS Code Copilot, Gemini CLI all speak MCP
- Every AI coding tool is becoming an agent runtime
- The team that owns the evaluation layer for MCP owns the reliability primitive for all of them

First-mover advantage is measured in months, not years.

Join us: make AI workflows reliable enough to ship

Jetty turns unreliable AI agents into production-grade workflows — by making evaluation a first-class primitive.

We're raising to:

- Scale evaluation infrastructure and trajectory storage
- Deepen enterprise compliance (SOC 2, HIPAA, on-premise)
- Capture the workflow layer before it consolidates

jetty.io · contact@jetty.io