

Total Cost of Ownership for Evaluating Agents

Fiddler Centor Models vs. LLMs at enterprise scale

Every observability platform promises visibility into your agents. But most don't tell you what it actually costs to evaluate them at scale. Your evaluation TCO is more than you may realize: there is a hidden cost called the Evaluation Trust Tax, incident risk exposure, and operational overhead.

How Platforms Evaluate Agents with LLMs

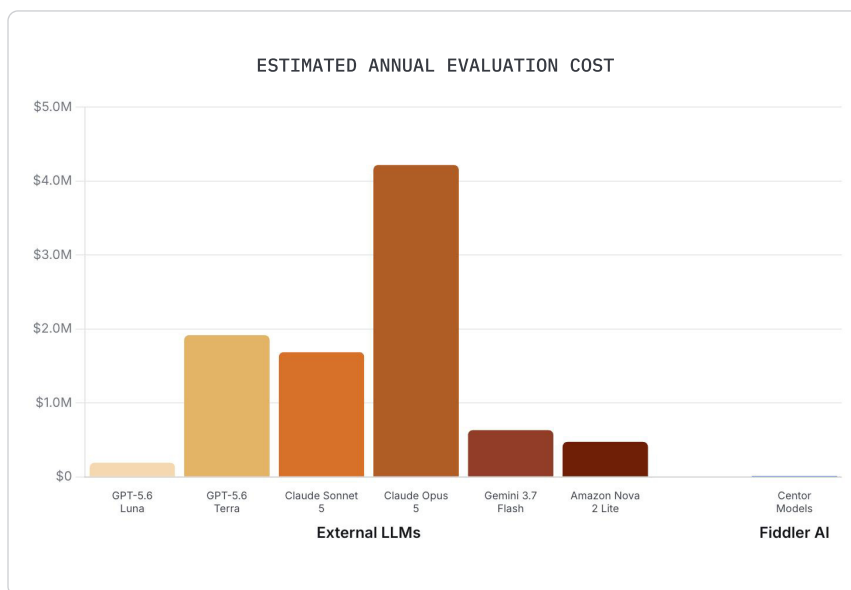
- 1 Your trace is generated by your agent
 - 2 The trace triggers API calls to an LLM provider for evaluation (OpenAI, Anthropic, etc.)
 - 3 The API costs show up on the bill from your LLM provider
- \$ You pay the Evaluation Trust Tax**

What Your Evaluation Cost Looks Like At Scale

Evaluation costs grow with trace volume, tokens per trace, and the number of evaluations per trace. Sampling plays a role too with LLMs. Centor Models do not sample and their costs are a step function compared to LLM costs that grow linearly. As your deployment sizes grow, the cost difference compounds at scale.

98% Cheaper

Fiddler Centor Models evaluate every trace, yet cost up to 98% less than external LLMs that sample 10% of the traces.



Calculate Your Evaluation TCO

Use this calculator to compare evaluation costs of Fiddler Centor Models vs. external LLMs, and see how quickly the hidden costs balloon.

fiddler.ai/evals-tco-calculator ↗

How External LLM Calls Drive Up Your Evaluation TCO



The Evaluation Trust Tax

You are charged every time a trace is evaluated via an external LLM call. This shows up on your LLM provider's bill. At enterprise scale, it compounds fast.



Incident Risk Exposure

To control costs, some teams may consider sampling, but the traces you skip could be the ones that matter most. These are low-frequency, high-impact events that sampling can miss.



Operational Overhead

Engineering time and effort should go to building agents, not maintaining evaluation infrastructure: API orchestration, model hosting, prompt versioning, and calibration.

Fiddler Centor Models For Evals and Policy Enforcement

Fiddler Centor Models (formerly known as Fiddler Trust Models) are cost-effective and secure models integral to the Fiddler AI Control Plane. They power the industry's fastest guardrails and evaluations, with no sampling, and no external LLM API costs. Whether evaluations require low-latency, task-specific evaluators or complex reasoning evaluators with high accuracy, Centor Models keep evaluation TCO low as agent traffic scales.

Teams can also bring their own judges (BYOJ) into Fiddler to use preferred LLMs alongside Centor Models, fitting the right evaluator to each agent and LLM project.

Centor Models Come In Two Types:



Out-of-the Box Models

- Hallucination detection, safety scoring, toxicity, jailbreak detection, and PII/PHI identification
- Ultra low-latency, cost-effective, and task-specific



Customizable Models

- Domain-specific prompt-based evaluators
- Built for complex, high accuracy reasoning

Trusted by Industry Leaders and Developers

"Fiddler delivered unified observability, protection, and governance across agents and predictive models, making it fundamental to our AI strategy."

Karthik Rao, CEO, Nielsen

◀▶ Nielsen

Thumbtack



tabs+



qualtrics^{XM}

OSCILAR

Fiddler is the AI Control Plane, the system of trust, for first-party and third-party agents from the creation layer to production. Continuous evaluation, reliable monitoring, enforceable policy, and auditable governance give enterprises the centralized controls and actionable insights to scale AI with trust. Integral to the platform are the secure Fiddler Centor Models, which power the industry's fastest guardrails and evaluations that require low latency and cost-effectiveness or complex reasoning with accuracy. Compared to external LLMs, Fiddler Centor Models offer a low TCO with no hidden costs.

Fortune 100 organizations use Fiddler to deliver high performance agentic and predictive applications, protect from costly risks, and maximize ROI.