

Fiddler Trust Service for LLM Observability and Security

Industry-leading guardrails – fast, cost-effective, and secure



AI Observability and Security is the foundation which gives enterprises the confidence to ship more predictive and generative models and LLM applications into production safely and responsibly.

Integral to the platform, the Fiddler Trust Service includes a series of purpose-built, fine-tuned Fiddler Trust Models that enable high quality LLM monitoring, scoring in live environments with the fastest guardrails in the industry. For added security, the Fiddler Trust Service can be deployed in cloud or VPC environments, ensuring enterprises maintain strict data control and safeguard LLM applications.

Fiddler Trust Models Deliver

<100ms

Guardrails Response Time*

6x

Cheaper**

5M+

Requests Per Day

Why Enterprise Leaders Choose the Fiddler Trust Service for LLM Monitoring

Fastest

With a <100ms latency, the models are optimized for rapid scoring, monitoring, and guardrails, ensuring enterprises can quickly detect and resolve LLM issues.

Secure

Fiddler can be deployed in cloud or VPC environments, maintaining compliance and protecting sensitive data.

Cost-Effective

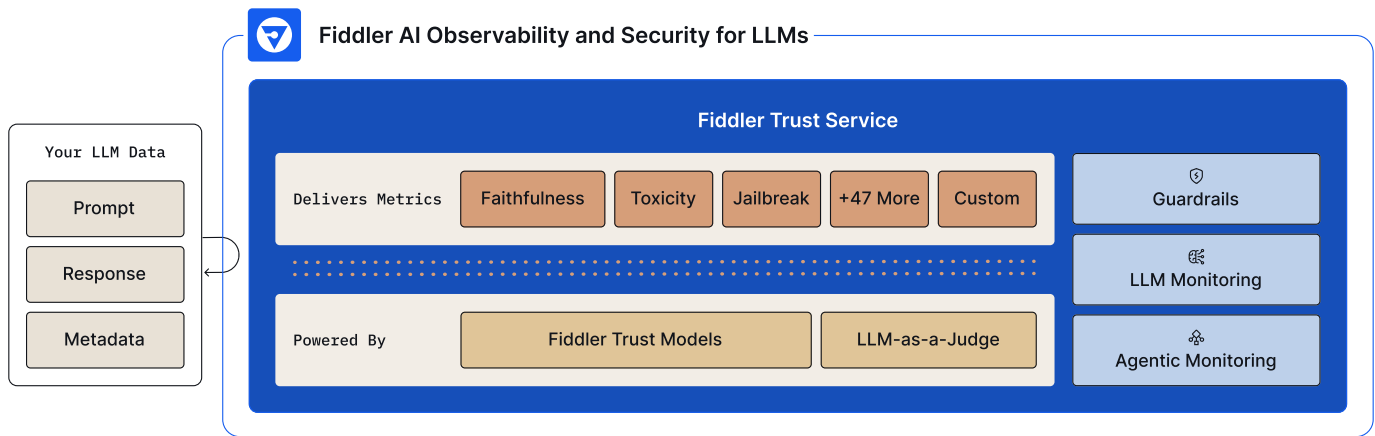
Trust Models are task-specific, optimized for efficiency and accuracy, while minimizing computational overhead.

Scalable

Built to support high-volume enterprise workloads, Fiddler Trust Service handles over 5 million requests per day out-of-the-box.

*Dependent on input size, geographical location, system load, or other infrastructure variability.

**Fiddler Trust Models are benchmarked against publicly available datasets.



Fiddler Trust Service: Quality and Moderation Controls for LLM Applications

The Fiddler Trust Service is an enterprise-grade solution that enables efficient use of computational resources and helps control costs compared to other LLM-as-a-judge offerings.

It combines two main scoring approaches: Fiddler Trust Models and LLM-as-a-judge. Together, they deliver 50 LLM metrics out-of-the box and custom metrics, powering guardrails, LLM monitoring, and agentic monitoring for secure enterprise deployments.

MODERATION CENTER

Guardrails

You, Fiddler can monitor ranking models. Fiddler enables monitoring for a ranking model by using a dataset from Expedia that includes shipping and purchase data with information on price competitiveness. The data is organized around a set of "search result impressions", which are the ordered list of results that the user sees after they search for a hotel on the Expedia website.

To get started with monitoring a ranking model using Fiddler, you can refer to the documentation provided in the notebook titled "Quickstart: Ranking Monitoring" on the Fiddler website. This notebook provides a detailed explanation of how Fiddler enables monitoring for a ranking model and includes instructions on how to use Google Cloud to get started with monitoring.

For more information on monitoring ranking models with Fiddler, you can visit the following URL: <https://docs.fiddler.ai/docs/ranking-models>

Additionally, you can refer to the code samples and instructions provided in the notebook to understand how to set up monitoring for ranking models using Fiddler.

If you have any further questions or need assistance with monitoring ranking models using Fiddler, you can join the Fiddler community Slack to ask questions or reach out to the support team for help.

Answer Faithfulness: 0.85 Jailbreak Likelihood: 0.001 Guardrails Latency: 184.2 ms

- Deploy the fastest guardrails securely at <100ms.
- Native to NVIDIA NeMo Guardrails.

QUALITY CENTER

LLM Monitoring

LLM Response Faithfulness Tracker

Faithfulness Outcomes

- Monitor hallucination, safety, and jailbreak attempts.
- Get real-time alerts on LLM issues.

Agentic Monitoring

- Gain deep visibility into AI agent behaviors.
- Interactive span-level drill-down analytics.

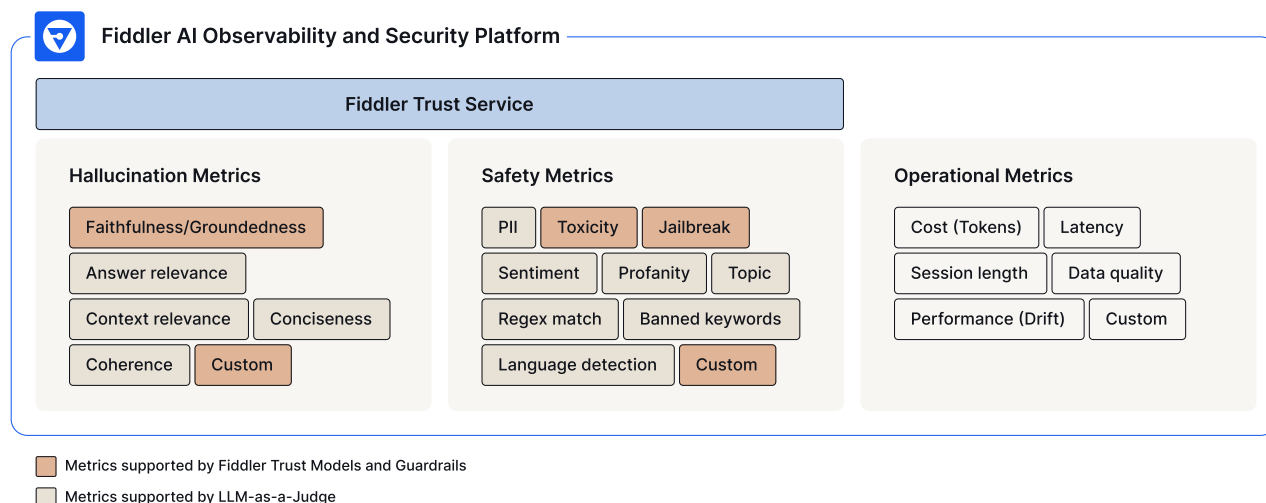
Fiddler Trust Models

In-house built models that provide fast, cost-effective, accurate and secure LLM scoring and metrics.

The Fiddler Trust Models are a series of proprietary, task specific models that score LLM prompts and responses based on hallucination, and toxicity, and jailbreak metrics. Compared to closed source foundation models, they provide highly accurate, cost effective, low latency scoring for a broad range of LLM metrics.

Enterprises can also create their own metrics by submitting prompts to the Fiddler-hosted Llama 3.1 8B model to address domain-specific use cases and specialized requirements. This fully-managed solution handles 300K+ daily events without the burden of infrastructure management.

Fast, Cost-Effective, and Secure Monitoring of 50 LLM Metrics



Fiddler Guardrails

LLM safety mechanism that moderates risky prompts and responses in real-time before they cause damage.

At <100ms latency, Fiddler Guardrails is the fastest in the industry. It leverages the scoring of the Fiddler Trust Models to evaluate prompts and responses and moderate harmful outputs for hallucination, toxicity, and jailbreaks. Simply specify your desired metric thresholds and let Fiddler Guardrails handle the enforcement.

The Fiddler Trust Service Excels at Popular and Niche Generative AI Use Cases



AI Chatbots

Boost investor value and confidence with accurate financial advice and recommendations from AI chatbots.



Internal Copilot Applications

Enhance employee productivity and boost their confidence in decision-making.



Compliance and Risk Management

Detect adversarial attacks and data leakage.



Content Summarization

Deliver highly accurate summaries for your users.



LLM Cost Management

Increase LLM operational efficiency gains.

Fiddler is the all-in-one AI Observability and Security platform for responsible AI. Monitoring and analytics capabilities provide a common language, centralized controls, and actionable insights to operationalize production ML models, GenAI, and LLM applications with trust. An integral part of the platform, the Fiddler Trust Service provides quality and moderation controls for LLM applications. Powered by cost-effective, task-specific, and scalable Fiddler-developed trust models — including air-gapped deployment for secure environments — it delivers the fastest guardrails in the industry.

Fortune 500 organizations use Fiddler to scale LLM and ML deployments to deliver high performance AI, reduce costs, and be responsible in governance.