



traceloop

Real-time monitoring and evaluation for LLM-based apps

Problem



No quantifiable metrics



LLMs are non-deterministic and can generate arbitrary content

Small changes to models and prompts can break existing behavior



No visibility into generated content



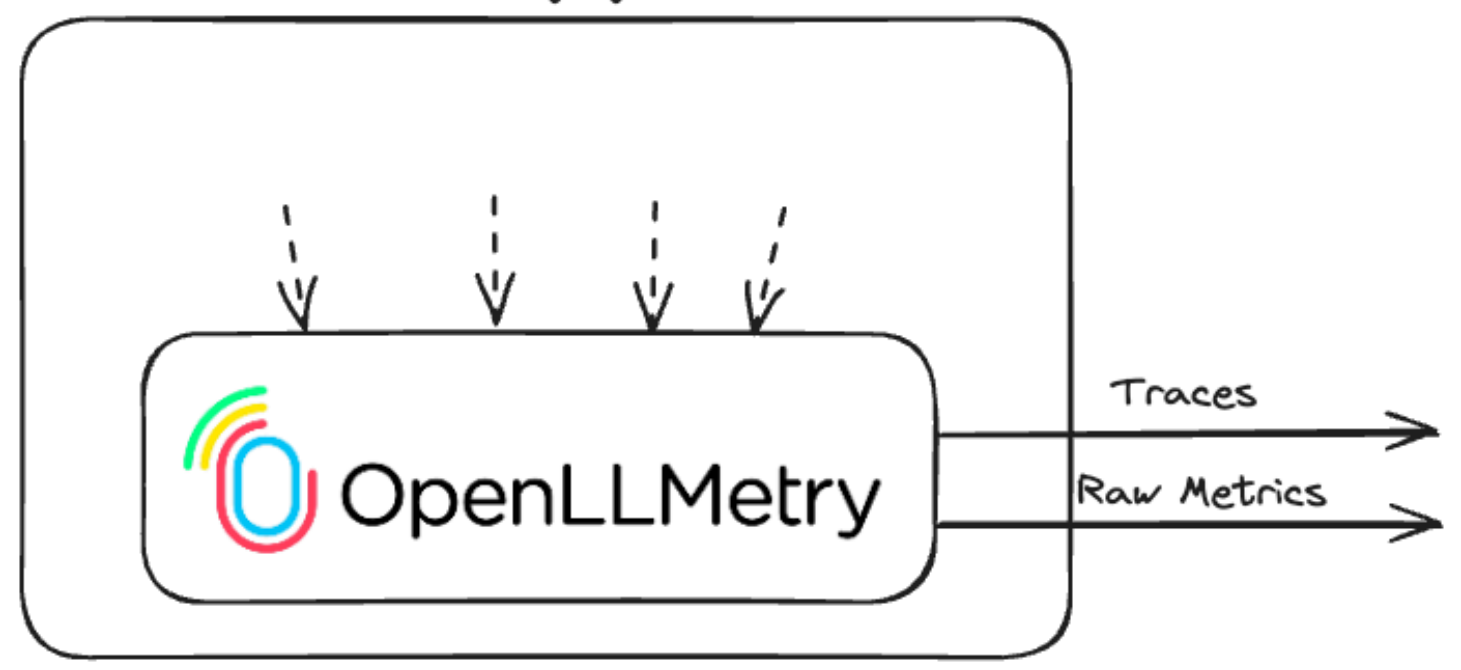
Solution

**Traceloop is a real-time complete
LLM observability with 1 line install**

- Get alerted on >90% of model hallucinations
- Track cost, usage and debug executions

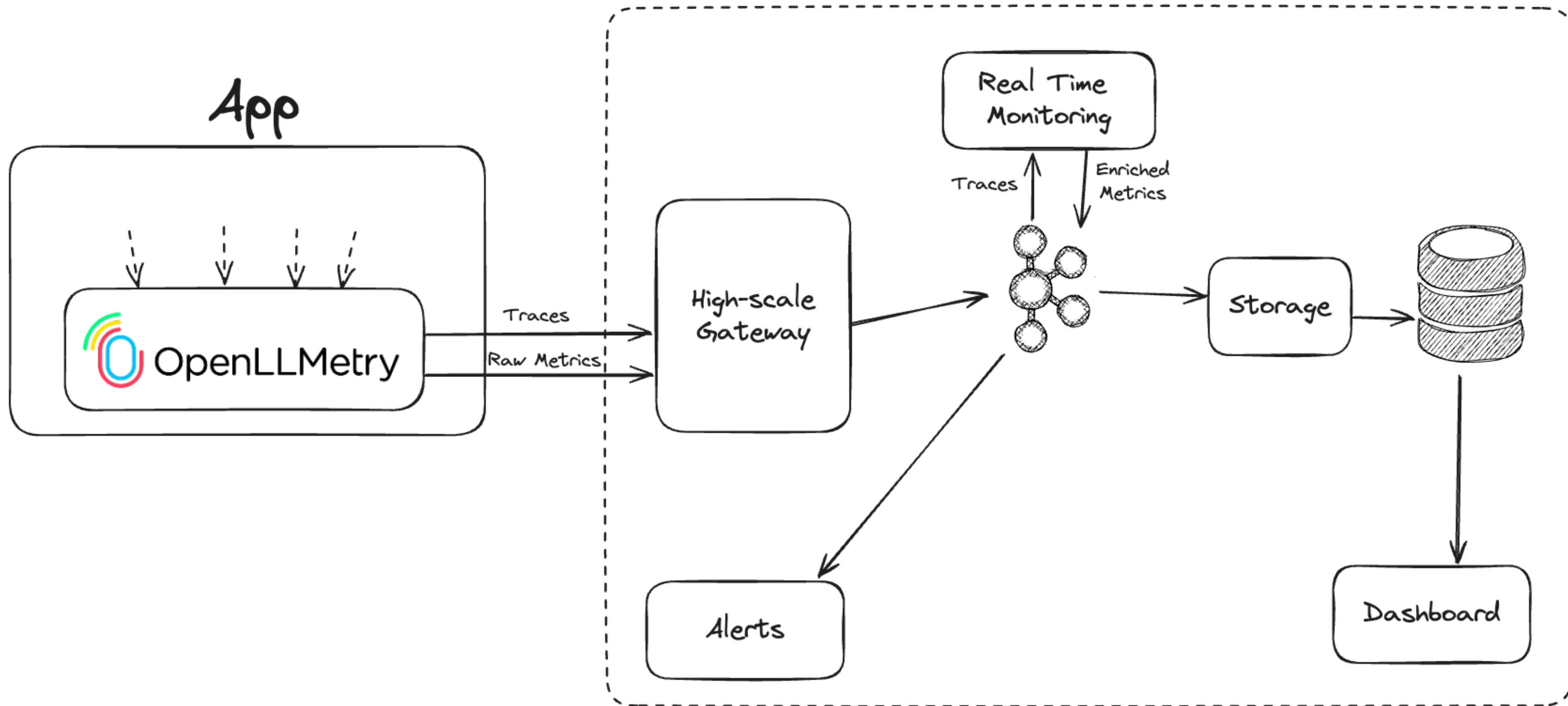


App



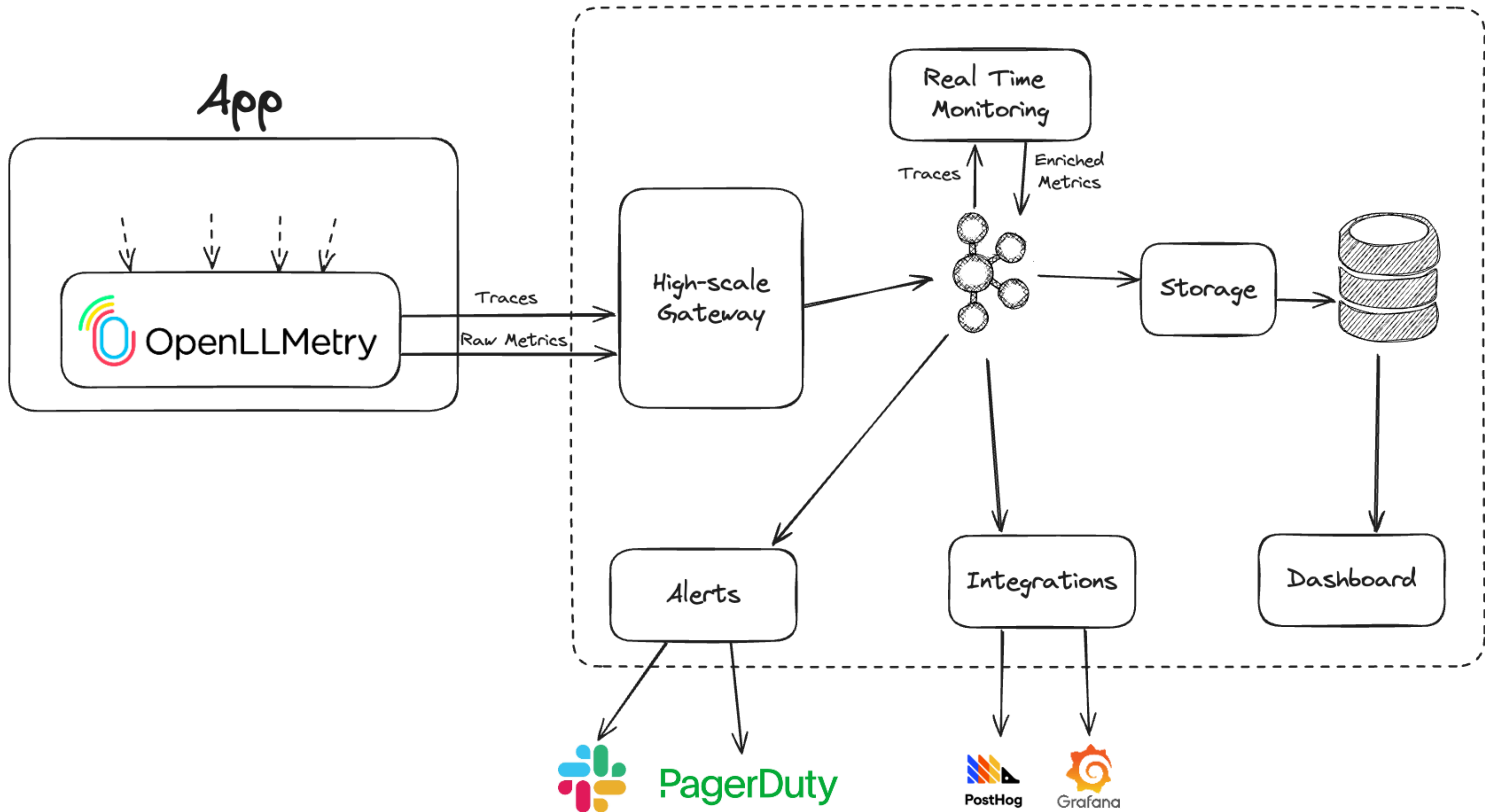


traceloop





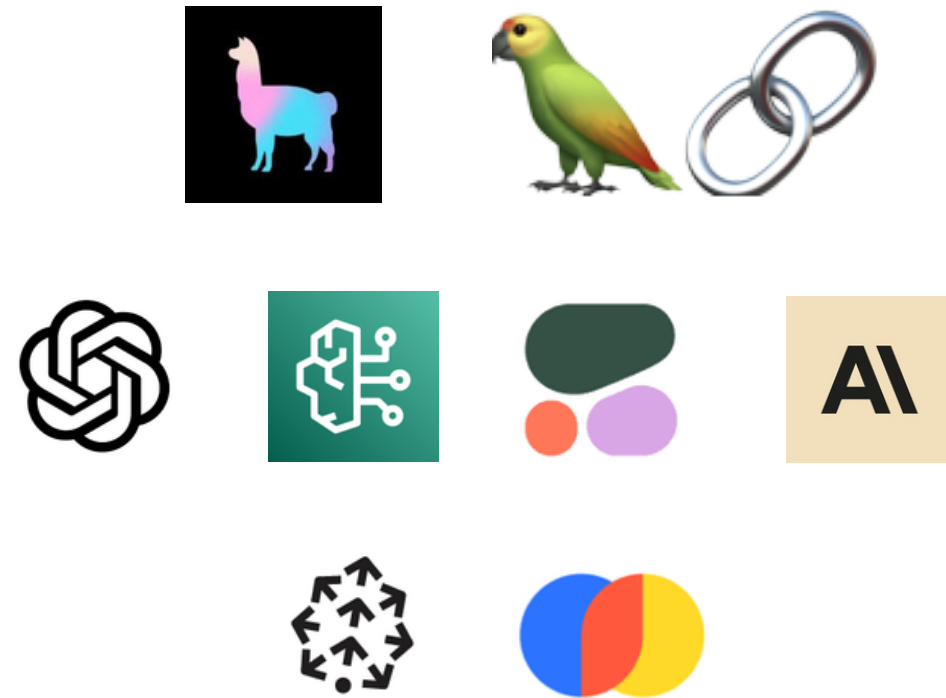
traceloop



OpenLLMetry

collaboration with OpenTelemetry

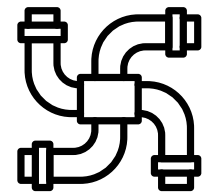
20+ providers
4 languages



The screenshot shows the Traceloop Trace details interface. The title is "Trace details" and the breadcrumb is "All traces > Support Respons". The interface displays a list of traces with columns for Name, Kind, and duration (0s, 0.6s, 1.2s, 1.8s, 2.4s). The traces are as follows:

Name	Kind	Duration
llama_index_retriever_query.workflow	Workflow	~2.4s
retrieve.task	Task	~0.6s
get_query_embedding.task	Task	~0.6s
chroma.query	Workflow	~1.2s
chroma.query.segment	Workflow	~1.2s
synthesize.task	Task	~2.4s
openai.chat	Workflow	~1.8s
openai.chat	Workflow	~0.6s

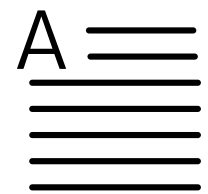
Metrics for every use case



RAG



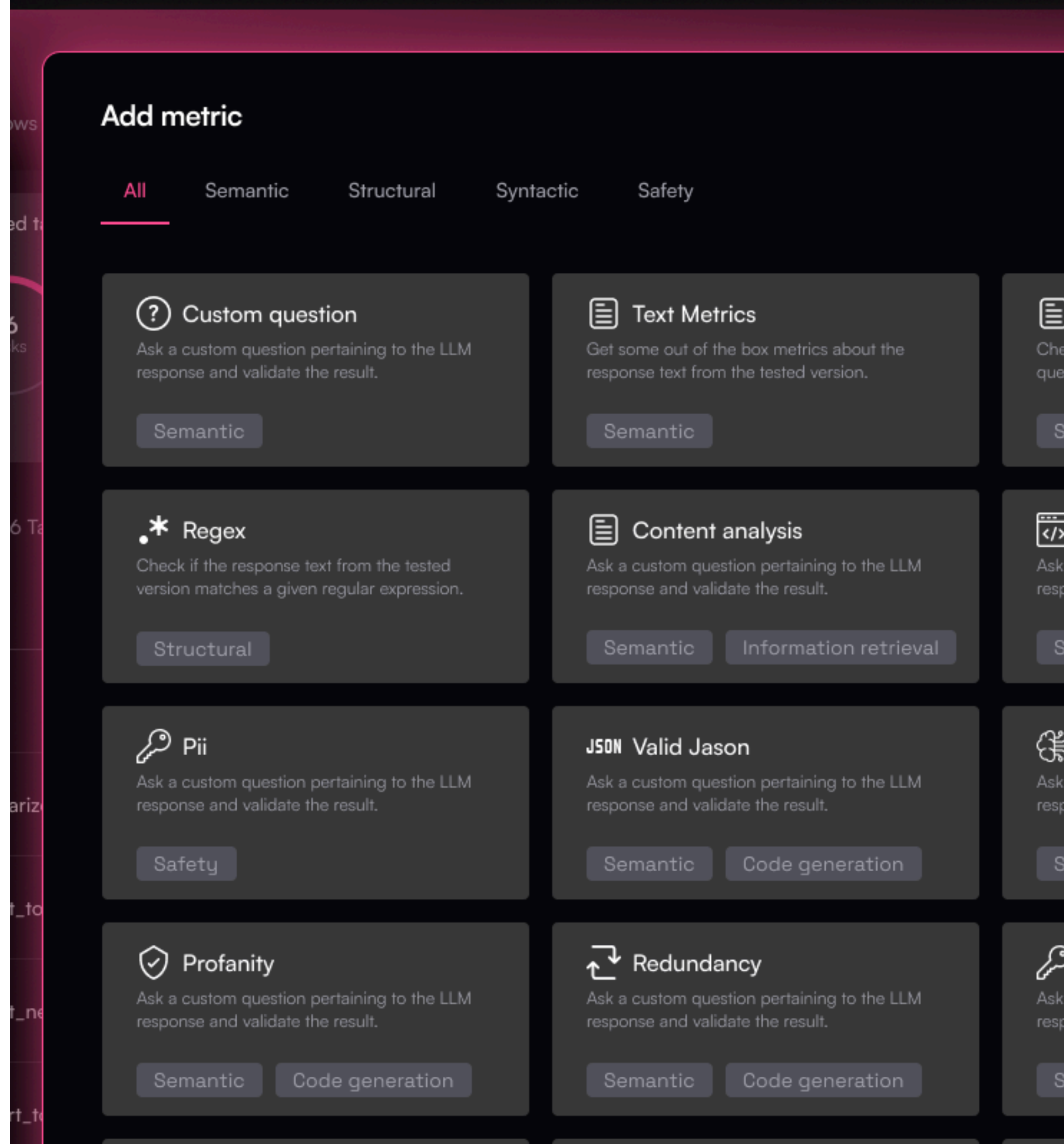
Entity Extraction



Summarization



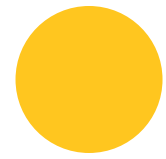
Chatbot



Real time monitoring

- High-throughput
- Low latency
- Correlates with human feedback





Pricing



Per-usage based

\$0.01

per trace

Volume discounts above 1M traces / month



Security



**On-prem & hybrid
deployment options**



Demo

