Comparison of Observability Platforms: LangSmith & Langfuse

June 2024

Iván Cordova
Charlie del Carpio

Comparison of Observability Platforms: LangSmith & Langfuse

AI solutions, particularly those leveraging Large Language Models (LLMs), have revolutionized various industries by providing advanced capabilities such as natural language understanding, text generation, and decision-making assistance. These LLMs, however, are often considered "black boxes" because they process user inputs and deliver predictions without transparent insight into their internal workings. This opacity can lead to challenges in understanding, debugging, and optimizing AI applications. This is where observability tools come into play.

Observability tools enable developers and stakeholders to monitor, analyze, and gain insights into the performance and behavior of AI models. By utilizing these tools, teams can track key metrics, identify anomalies, and ensure the models are operating as intended. In essence, observability tools transform the black box nature of LLMs into a more transparent and manageable system, fostering trust and reliability in AI-driven solutions. This transparency is crucial for maintaining the effectiveness and accuracy of AI applications, ultimately leading to better user experiences and more robust AI systems.

LangSmith and Langfuse are leading observability platforms for AI applications, offering comprehensive monitoring and analytics tools. Both platforms enable you to achieve excellent results in tracking and optimizing model performance. However, if you seek advanced capabilities, such as receiving alerts when certain criteria are met, LangSmith is the best option. Langfuse, being open source, remains a flexible choice with an active roadmap to incorporate many of LangSmith's unique features in the future.

The features evaluated for the comparison matrix are the following

Tracing

Tracking the flow of data and operations through an AI system to diagnose issues, optimize performance, and ensure transparency. It helps in understanding how inputs are processed to produce outputs.

Human Feedback

Incorporating insights and inputs from users/domain experts to refine and enhance AI models. This feedback loop is crucial for improving model performance and ensuring it meets real-world requirements.

Dataset Management

Encompasses the practices and tools used to organize, store, and maintain datasets. Effective dataset management ensures data quality, facilitates easy access, and supports robust AI model training and evaluation.

Monitoring & Automation

Continuously overseeing AI systems to ensure they operate correctly and efficiently. Automated tools can alert teams to issues, perform routine maintenance, and optimize system performance without manual intervention.

Evaluation

Assessing the performance and accuracy of AI models using various metrics and benchmarks. This process helps in identifying strengths, weaknesses, and areas for improvement in AI applications.

Prompt Management

Creating, organizing, and optimizing prompts used to interact with AI models, particularly in natural language processing tasks. Effective prompt management ensures models understand and respond accurately to user inputs.

Both LangSmith and Langfuse can be self-hosted in your own cloud environment, catering to teams with heightened security, deployment, and support needs. Langfuse offers this capability for free, but its enterprise plan provides additional features not available in the open-source version, such as the LLM Playground and other advanced tools. On the other hand, LangSmith's self-hosting option is part of its enterprise plan, which requires an annual payment. This plan includes all features from the developer version and adds significant benefits like enhanced deployment support, custom rate limits, training, architectural guidance, and more. Self-hosting these applications ensures that you have greater control and security over your AI observability tools while benefiting from the comprehensive support offered by these platforms.

Typically, LangSmith is used via the provided platform, whereas Langfuse is often self-hosted or locally hosted. For additional features in Langfuse, users can opt for the team (enterprise) edition. Similarly, to access more benefits in LangSmith, users need to subscribe to the enterprise plan.

Comparison of Observability Platforms

LangSmith

Langfuse

Pros

- Extensive feature set
- Robust tracing
- Customizable evaluations
- Strong dataset management
- Easy configuration

- Low performance overhead
- Built for complex use cases
- Comprehensive analytics
- Extensive integration ecosystem: Langchain, LlamaIndex, and other frameworks through SDKs

Cons

- Higher pricing
- No documentation for integration with frameworks other than LangChain.

- May lack advanced features like multimodal tracing and triggers found in LangSmith
- Potential learning curve for new users because it is developer oriented

Tracing

- Detailed and custom function tracing with '@traceable' decorator
- Distributed tracing with context propagation
- Multimodal tracing

- Full context capture
- Session grouping for holistic views
- Optimized for low performance overhead
- Trace complex LLM applications with "@observe" decorator

Dataset Management

- Comprehensive data management: support the creation, management, and updating of datasets from various sources

- Comprehensive data management: support the creation, management, and updating of datasets from various sources

Evaluation

- Custom evaluators and metrics
- Pairwise and regression testing
- Real-time and historical evaluations
- Heuristic and LLM-as-judge evaluators

- Custom evaluators and metrics
- Real-time and historical evaluations
- Integration and automation

Human Feedback

- Integration of human feedback
- Support for human annotation

- Integration of human feedback
- Support for human annotation

Monitoring & Automation

- Interactive monitoring dashboards
- Real time feedback and auto evaluation
- Automation rules for proactive monitoring and webhooks to trigger actions over events related to the LLM behavior

- Interactive monitoring dashboards
- Real time feedback and auto evaluation

Prompt Management

- Centralized prompt management and versioning
- Integration for Langchain Hub

- Centralized prompt management and versioning

Astral Insights is a boutique AI enablement firm specializing in deploying scalable AI platforms and assistants to optimize workflows across the supply chain. From education and consulting to design and implementation, we are a one-stop-shop to bolster competitive advantage. Our customized systems automate tasks, optimize workflows, and spur innovation for efficiency and profitability. Guided by our core values of accountability, creativity and excellence, we help our clients integrate AI-enabled solutions to achieve tangible results.

Sources:
LangSmith Official Documentation - Comprehensive features including tracing, dataset management, evaluation, and monitoring: https://docs.smith.langchain.com

Langfuse Features Overview - Details on context capture, session grouping, prompt management, and analytics: https://langfuse.com