

Category

Tracing SDKs

Examples

OpenTelemetry, LangSmith, Arize Phoenix

Strength

Capture granular interaction data (prompts, responses, latency, token counts)

Limitation

Require engineering work to operationalize; no built-in governance workflow

Evaluation platforms

Braintrust, Vellum, DeepEval

Score outputs against quality and compliance criteria; support human review workflows

Typically standalone; require integration with governance and alerting infrastructure

Governance platforms

Dataiku Govern, IBM watsonx governance

Lifecycle management, approval workflows, audit trails, compliance reporting

Platform-level investment; strongest value at enterprise scale