**Requirement:-**

Need a tool that supports evals and guardrails, with metric tracking and model lifecycle management.

**Tools Considered:-**

MLflow

DeepEval

NVIDIA NeMo Guardrails

LangChain

Guardrails AI

**Tool evaluation Metrix**

| **Criteria** | **MLflow + DeepEval** | **NVIDIA NeMo**  **Guardrails** | **LangChain** | **Guardrails AI** |
| --- | --- | --- | --- | --- |
| Evaluation Metric | Advanced | Limited | Basic | Basic |
| Model Tracking and Versioning | Advanced | None | None | None |
| Guardrails flexibility | Customizable | Rule-based | Customizable | Limited |
| Deployment Monitoring | Supported | None | Basic | None |
| Visualization | Rich dashboard | Limited | Limited | Limited |

**Recommendation:** MLflow + DeepEval

**References**:

<https://mlflow.org/docs/latest/llms/llm-evaluate/index.html>

<https://www.confident-ai.com/blog/llm-guardrails-the-ultimate-guide-to-safeguard-llm-systems>