

Aplicaciones LLM

Precaución: Cuestiones Delicadas

LangSmith nos ayuda con:

- Visualización y testing de componentes.
- Visualización y formateado de datos de contexto.
- Visualización de la ingeniería de prompts.
- Visualización del output con Foundation LLM alternativos.
- Evaluación: calidad del output de la aplicación.
- Colaboración: acceso del equipo al proceso de debugging.

Visualización y testing de componentes

- Visualizar los componentes de la Aplicación LLM.
- Experimentar fácilmente visualizando diferentes enfoques.
- Testing y debugging componentes.

Visualización y formateado de datos de contexto

- Testing y debugging de los datos utilizados para el contexto.
- Observability: qué datos exactos se pasan para el contexto.

Visualización de la ingeniería de prompts

- Cómo se ve exactamente el prompt cuando entra en el LLM.
- Cómo se ve el prompt después de varios loops y operaciones que le pueden afectar.
- Testing y debugging prompts.

Visualización del output con Foundation LLM alternativos

- Probar la Aplicación LLM con diferentes Foundation LLM.

Evaluación: calidad del output de la aplicación

- Los dos problemas principales al evaluar Aplicaciones LLM son la falta de datos y la falta de buenas métricas.
 - A diferencia del ML tradicional no tienes una dataset de prueba inicial. Puedes tener un MVP sin tener un conjunto de datos de prueba inicial. Es una ventaja, pero también un desafío para la evaluación.
- Cómo solucionar la falta de datos. Es bueno construir una dataset de evaluación: manualmente, sintéticamente (con un LLM) o a partir de feedback de usuarios.
- Cómo solucionar la falta de métricas. Las métricas cuantitativas no funcionan muy bien. Mucha gente todavía hace algo así como una "vibes check", construye intuición. LangSmith facilita la construcción de intuición mostrando las entradas y outputs del Foundation LLM.
- Evaluación asistida por LLM: un LLM evalúa las salidas de la Aplicación LLM.
- Llevar un registro del feedback de los usuarios ayuda con la evaluación.
- Hacer AB testing no es tan fácil como con aplicaciones tradicionales, pero puede ser útil.

Colaboración: acceso del equipo al proceso de debugging

- LangSmith permite el acceso en equipo al proceso de debugging.
- Los datos de contexto pueden ser gestionados por Científicos de Datos.
- El rendimiento de la aplicación puede ser gestionado por Product Managers.
- El desarrollo de la aplicación puede ser gestionado por el Ingeniero de la Aplicación LLM.