

LLM Apps

Debugging with LangSmith

© 2023 Julio Colomer, AI Accelera

LangSmith helps us with:

- Component visualization and testing.
- Context data visualization and formatting.
- Prompt engineering visualization.
- Visualize output with alternative Foundation Models.
- Evaluation: quality of the output of the app.
- Collaboration: team access to the debugging process.

Component visualization and testing

- Visualize the components of the LLM App.
- Experiment easily visualizing different approaches.
- Test and debug components.

Context data visualization and formatting

- Test and debug the data used for context.
- Observability: what exact data is passed for context.

Prompt engineering visualization

- How exactly looks the prompt when it goes into the LLM.
- How the prompt looks like after several loops and operations that may affected it.
- Test and debug prompts.

Visualize output with alternative Foundation Models

- Test how the LLM App does with different Foundation Models.

Evaluation: quality of the output of the app

- 2 main problems evaluating LLM Apps: lack of data, lack of good metrics.
 - Contrary to traditional ML, you do not have an initial test dataset. You can have an MVP without having an initial test dataset. That is awesome, but it is a challenge for evaluation.
- How to solve lack of data. It is good to build evaluation datasets: manually, synthetically (with an LLM) or from user feedback.
- How to solve lack of metrics. Quantitative metrics do not work very well. Many people still does kind of a “vibe check”, builds intuition. LangSmith makes easy to build intuition by showing the inputs and the outputs of the Foundation Model.
- LLM-assisted evaluation: an LLM evaluates the outputs of the LLM App.
- Keeping track of user feedback helps with evaluation.
- Doing AB testing is not as easy as with traditional apps, but can be helpful.

Collaboration: team access to the debugging process

- LangSmith allows team access to the debugging process.
- Context data is best handled by Data Scientists.
- App performance is best handled by Product Managers.
- App development is best handled by the LLM App Engineer.