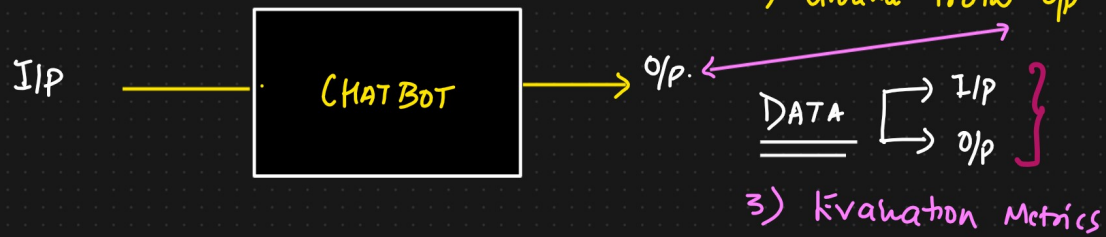


Evaluation OF RAG Application AND CHATBOT



1) LLM AS A Judge + Prompt

LangSmith

1) Gathering DATA points

2) LLM AS A JUDGE

3) Evaluation Metrics

4) Comparison Multiple LLM models.

② RAG EVALUATION

1) How to create test datasets?

LangSmith ✓

2) RUN RAG APP → Test DATASETS

3) Measure RAG performance → Different Evaluation Metrics.

Steps

Experiments

① RAG → DATA Ingestion, Retriever, Generation ✓

② TEST DATA ⇒ Question ↔ Answer

③ Evaluation Metrics [LLM as a judge]

