

Q121. Which tool is primarily used for labeling, feedback collection, and evaluation dashboards?

- A. LangSmith
 - B. Argilla ☒
 - C. Trulens
 - D. PromptLayer
-

Q122. Which of the following tools evaluates RAG quality metrics such as faithfulness and relevance?

- A. Trulens ☒
 - B. Weights & Biases
 - C. LangSmith
 - D. HumanEval
-

Q123. PromptLayer is mainly used for:

- A. Debugging vector databases
 - B. Monitoring and logging prompts ☒
 - C. Evaluating relevance scores
 - D. Tracking user feedback
-

Q124. Which tool is widely used to track experiments and monitor RAG performance?

- A. LangSmith
 - B. Trulens
 - C. Weights & Biases (W&B) ☒
 - D. HumanEval
-

Q125. LangChain's debugging + tracing + evaluation tool is called:

- A. Argilla
 - B. LangSmith ☒
 - C. PromptLayer
 - D. HumanEval
-

Q126. Which tool is most suitable when you want manual annotation for evaluation?

- A. HumanEval ☒
 - B. PromptLayer
 - C. LangSmith
 - D. W&B
-

Q127. For collecting user feedback and labeling datasets in RAG pipelines, the best tool is:

- A. PromptLayer
 - B. Argilla ☒
 - C. LangSmith
 - D. Trulens
-

Q128. Which RAG evaluation tool focuses on measuring faithfulness (truthfulness) of responses?

- A. Argilla
 - B. Trulens ☒
 - C. PromptLayer
 - D. W&B
-

Q129. If you want to monitor prompts and their performance over time, which tool should you choose?

- A. PromptLayer ☒
 - B. LangSmith
 - C. Argilla
 - D. HumanEval
-

Q130. To debug retrieval mismatches and trace document relevance in LangChain, use:

- A. LangSmith ☒
 - B. Argilla
 - C. W&B
 - D. PromptLayer
-

Q131. Which tool integrates feedback loops directly into RAG pipelines?

- A. Argilla ☒
 - B. PromptLayer
 - C. HumanEval
 - D. Trulens
-

Q132. For tracking hyperparameters and experiment results in RAG, use:

- A. PromptLayer
 - B. Weights & Biases ☒
 - C. Argilla
 - D. LangSmith
-

Q133. Which RAG evaluation tool provides dashboards for annotators and reviewers?

- A. Argilla ☒
 - B. Trulens
 - C. HumanEval
 - D. PromptLayer
-

Q134. To evaluate hallucinations and irrelevant responses in RAG, you would prefer:

- A. Trulens ☒
 - B. Argilla
 - C. W&B
 - D. LangSmith
-

Q135. HumanEval is most useful for:

- A. Automating document retrieval
 - B. Manual annotation and dataset scoring ☒
 - C. Managing embeddings
 - D. Logging API keys
-

Q136. Which of the following tools does NOT directly provide evaluation capabilities?

- A. PromptLayer ☒
 - B. Trulens
 - C. Argilla
 - D. LangSmith
-

Q137. Which tool works best with LangChain for tracing RAG workflows?

- A. W&B
 - B. LangSmith ☒
 - C. Trulens
 - D. Argilla
-

Q138. If you want to combine human feedback with automated evaluation, which tool is preferred?

- A. Argilla ☒
 - B. PromptLayer
 - C. W&B
 - D. HumanEval
-

Q139. Which tool provides metrics dashboards to monitor RAG pipelines in real-time?

- A. W&B ☒
 - B. LangSmith
 - C. Argilla
 - D. Trulens
-

Q140. For debugging vector store retrieval mismatches, the most recommended tool is:

- A. LangSmith ☒
 - B. PromptLayer
 - C. Trulens
 - D. HumanEval
-

Q141. Which tool allows you to log every prompt and response sent to an LLM?

- A. PromptLayer ☒
 - B. Argilla
 - C. Trulens
 - D. W&B
-

Q142. If you want to manually annotate relevance scores, you would choose:

- A. HumanEval ☒
 - B. PromptLayer
 - C. W&B
 - D. LangSmith
-

Q143. Which tool has built-in faithfulness and relevance metrics for RAG responses?

- A. Trulens ☒
 - B. LangSmith
 - C. Argilla
 - D. PromptLayer
-

Q144. To compare different RAG model performances visually, the best choice is:

- A. Weights & Biases ☒
 - B. PromptLayer
 - C. Trulens
 - D. HumanEval
-

Q145. Which of these tools helps you trace step-by-step RAG workflows inside LangChain apps?

A. LangSmith ☒

B. Trulens

C. Argilla

D. PromptLayer