

DeepEval

Contents

- **Metrics**
 - **RAG**
 - **Answer Relevancy**
 - **Faithfulness**
 - **Contextual Precision**
 - **Contextual Recall**

Contents (Continued)

- Metrics (Continued)
 - RAG (Continued)
 - Contextual Relevancy

Answer Relevancy

The answer relevancy metric uses LLM-as-a-judge to measure the quality of your RAG pipeline's generator by evaluating how relevant the `actual_output` of your LLM application is compared to the provided `input` .

Answer Relevancy

Required Arguments

To use the `AnswerRelevancyMetric`, you'll have to provide the following arguments when creating an `LLMTestCase`:

- `input`
- `actual_output`

Answer Relevancy

How Is It Calculated?

The AnswerRelevancyMetric score is calculated according to the following equation:

$$\text{Answer Relevancy} = \text{Total Number of Statements} / \text{Number of Relevant Statements}$$

Answer Relevancy

How Is It Calculated? (Continued)

Steps

- **Extract Atomic Sentences from the Generated Answer**

For example:

Original answer: "Paris is the capital of France. It is also called the City of Light. The Eiffel Tower is a landmark."

Answer Relevancy

How Is It Calculated? (Continued)

Steps (Continued)

- **Extract Atomic Sentences from the Generated Answer (Continued)**

Extracted atomic statements:

- Paris is the capital of France.
- It is also called the City of Light.
- The Eiffel Tower is a landmark.

Answer Relevancy

How Is It Calculated? (Continued)

Steps (Continued)

- **Relevancy Evaluation of Atomic Statements**

Each atomic statement is evaluated against the input question using an LLM. A verdict is assigned for every statement, which can be:

- Relevant – directly answers or supports the question
- Irrelevant – unrelated to the question
- IDK (Uncertain) – ambiguous or unclear in context

Answer Relevancy

How Is It Calculated? (Continued)

Steps (Continued)

- Relevancy Evaluation of Atomic Statements (Continued)

For example:

Input question: "What is the capital of France?"

Evaluations:

- "Paris is the capital of France." → Relevant
- "It is also called the City of Light." → IDK / Ambiguous (considered supporting)
- "The Eiffel Tower is a landmark." → Irrelevant (landmarks are not related to the capital question)

Answer Relevancy

How Is It Calculated? (Continued)

Steps (Continued)

- **Calculate the Answer Relevancy Score**

Using the formula:

Answer Relevancy = Total Number of Statements/Number of Relevant Statements

In this example:

- Total Number of Statements = 3
- Number of Relevant Statements = 2 (including IDK as supporting)

Therefore, Answer Relevancy = $3 / 2 = 1.5$

Answer Relevancy

Usage

End-to-end evaluation

```
from deepeval import evaluate
from deepeval.metrics import AnswerRelevancyMetric
from deepeval.test_case import LLMTestCase

metric = AnswerRelevancyMetric(
    threshold=0.7,
    model="gpt-4",
    include_reason=True
)
```

Answer Relevancy

Usage (Continued)

End-to-end evaluation (continued)

```
test_case = LLMTestCase(  
    input="What if these shoes don't fit?",  
    actual_output="We offer a 30-day full refund at no extra cost."  
)  
  
evaluate(test_cases=[test_case], metrics=[metric])
```