RAGAS: Answer Relevancy

What is it?

Measures if the chatbot's answer actually addresses the user's question.
Score: 0-1 (higher = better)

How it Works

1. LLM generates 3 questions from the answer
2. Compares with original question using cosine similarity
3. Averages scores = final score

What You Provide

user_input: The question asked
response: Chatbot's answer

RAGAS handles everything else internally.

Code Example

```
from ragas.metrics import AnswerRelevancy
from ragas.llms import llm_factory
from openai import AsyncOpenAI

client = AsyncOpenAI()
llm = llm_factory("gpt-4o-mini", client=client)
scorer = AnswerRelevancy(llm=llm)

result = await scorer.ascore(
    user_input="What is your refund policy?",
    response="Refunds available within 30 days."
)
print(result)
```

Example 1: HIGH Score (Good Answer)

Question: How can I cancel my flight?

Answer: Cancel via Flynas app under My Trips section.

Generated Qs: How to cancel flight? Where to cancel booking?

Result: Questions MATCH original → Score: 0.92 (HIGH)

Example 2: LOW Score (Off-topic Answer)

Question: How can I cancel my flight?

Answer: Flynas offers delicious meals on all flights.

Generated Qs: What meals offered? How to order food?

Result: Questions DON'T match → Score: 0.15 (LOW)

Key Insight

If the answer is relevant, generated questions will match the original question. Off-topic answers produce different questions = low score.

Reference

https://docs.ragas.io/en/stable/concepts/metrics/available_metrics/answer_relevance/#how-its-calculated