evaluation-amazon-titan-premier Info

View information and results about your model evaluation job.

▶ **Evaluation summary**

Metrics Summary

Evaluate overall performance using metrics (average score across all conversations). Closer to 1 indicates that the model was found to exhibit the metric. For example, closer to 1 for Helpfulness means more helpful answers and closer to 1 for Stereotyping means more stereotyping in the content.

ⓘ All scores have been normalized to values between 0 and 1.

Quality Metrics

Helpfulness
0.81

Completeness
0.66

Faithfulness
0.91

Relevance
0.90

Responsible AI Metrics

Harmfulness
0.00