


LLM Comparison Framework


Upload your dataset with 'text' and 'reference_summary' columns.

Choose a CSV file



Drag and drop file here
Limit 200MB per file • CSV

Browse files

test_data.csv 360.0B

×

Dataset loaded successfully.

	text	reference_summary
0	OpenAI has announced a new AI model that performs natural language processing ta	OpenAI introduces a new
1	NASA's Perseverance rover has collected new samples that provide insights into the p	NASA's rover collects san

Enter model names to compare, separated by commas.

Model names

facebook/bart-large-cnn, google/pegasus-xsum, bedrock:amazon.titan-text-lite-v1

Initializing framework...

Initializing the framework with models and dataset.

Initializing AWS Bedrock client.

Evaluate Models

Evaluating all models on the dataset.

Evaluating model: facebook/bart-large-cnn

Loading model: facebook/bart-large-cnn

Generating summary for a given text.

Calculating ROUGE scores.

Calculating BERT-Score.

Generating summary for a given text.

Calculating ROUGE scores.

Calculating BERT-Score.

Evaluating model: google/pegasus-xsum

Loading model: google/pegasus-xsum

Generating summary for a given text.

Calculating ROUGE scores.

Calculating BERT-Score.

Generating summary for a given text.

Calculating ROUGE scores.

Calculating BERT-Score.

Evaluating model: bedrock:amazon.titan-text-lite-v1

Loading model: bedrock:amazon.titan-text-lite-v1

Generating summary for a given text.

bedrock model id : bedrock:amazon.titan-text-lite-v1

Generating summary using Amazon Bedrock model: bedrock:amazon.titan-text-lite-v1

Successfully generated summary using Bedrock.

Calculating ROUGE scores.

Calculating BERT-Score.

Generating summary for a given text.

bedrock model id : bedrock:amazon.titan-text-lite-v1

Generating summary using Amazon Bedrock model: bedrock:amazon.titan-text-lite-v1

Successfully generated summary using Bedrock.

Calculating ROUGE scores.

Calculating BERT-Score.

Visualizing ROUGE scores for all models.

	Model	ROUGE-1	ROUGE-2	ROUGE-L	BERT-Score
0	facebook/bart-large-cnn	0.2682	0.0866	0.2346	0.9079
1	google/pegasus-xsum	0.254	0	0.2169	0.8951
2	bedrock:amazon.titan-text-lite-v1	0.1871	0.0791	0.1497	0.8396