Model Comparison Report

Model Comparison Based on Metrics

Y-axis: Metric Value
X-axis: Models and Metrics

Bars: gpt-3 (BLUE Score), gpt-4 (BLUE Score), gpt-3 (Accuracy), gpt-4 (Accuracy)

Likert Scale Comparison of Models

Y-axis: Count
X-axis: Models and Metrics

Bars: gpt-3 (Usability Gap), gpt-4 (Cultural Gap), gpt-3 (Cultural Gap), gpt-4 (Usability Gap)

Legend: Inaccurate, Partially Accurate, Accurate