

Model Performance Comparison

BERTScore

- QwenVL2.5-72B
- Gemini-2.5-Flash
- GPT-4o
- InternVL3-78B
- Overall

OE-Eval

ROUGE

BLEU

METEOR

