

# KLSBench: Model Performance Across All Tasks

- EXAONE-3.0-7.8B-Instruct
- Qwen2.5-7B-Instruct
- claude-3-opus-20240229
- claude-sonnet-4-5-20250929
- gpt-3.5-turbo
- gpt-4-turbo
- Llama-3.1-8B-Instruct

Classification

1.0

0.8

0.6

0.4

0.2

Translation

NLI

Punctuation

Retrieval

