Length of LLM Answers vs. Expert Answer Lengths (Longest Samples)

Model GPT-4o — Correlation: 0.01

Model GPT-4o-mini — Correlation: 0.09

Model LLAMA3.1 8b — Correlation: -0.05

Model Open Mixtral 8x22b — Correlation: 0.09

Model Mistral-largest-latest — Correlation: 0.10

Model Gemma2 — Correlation: -0.02