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

Model GPT-4o — Correlation: 0.14

Model GPT-4o-mini — Correlation: 0.23

Model LLAMA3.1 8b — Correlation: 0.03

Model Open Mixtral 8x22b — Correlation: 0.31

Model Mistral-largest-latest — Correlation: 0.32

Model Gemma2 — Correlation: 0.03

(All plots: x-axis "Expert Answer Length", y-axis "LLM Answer Length")