

Correlation Statistics for Model–Dataset Pairs

1 Correlation between token entropy and semantic similarity of tokens

The entropy of token probability distributions has been proposed as a measure for LLM uncertainty. We propose a new uncertainty measure, which is the average pairwise cosine similarity of the top-p tokens of a probability distribution. In the following, we will show that there is a inverse relationship

Table: Pearson and Spearman Correlations

Model	Dataset	Pearson r	Pearson p	Spearman r	Spearman p
[t]3*Qwen3-8B	gsm8k	-0.37	0.00	-0.39	0.00
	writingprompts	-0.38	0.00	-0.41	0.00
	xsum	-0.37	0.00	-0.43	0.00
[t]3*Mistral-7B-v0.1	gsm8k	-0.50	0.00	-0.67	0.00
	writingprompts	-0.49	0.00	-0.66	0.00
	xsum	-0.51	0.00	-0.67	0.00
[t]3*Llama-3.1-8B-Instruct	gsm8k	-0.57	0.00	-0.70	0.00
	writingprompts	-0.58	0.00	-0.70	0.00
	xsum	-0.57	0.00	-0.69	0.00
[t]3*Llama-3.1-8B	gsm8k	-0.62	0.00	-0.76	0.00
	writingprompts	-0.61	0.00	-0.75	0.00
	xsum	-0.63	0.00	-0.76	0.00
[t]3*deepseek-llm-7b-base	gsm8k	-0.42	0.00	-0.65	0.00
	writingprompts	-0.43	0.00	-0.67	0.00
	xsum	-0.45	0.00	-0.67	0.00