LLMs in CT24 dataset

Partition	Model	Accuracy	Precision	f1	Recall
Test	Llama-2-7b	0.9091	0.8202	0.8249	0.8295
	GPT-2	0.8622	0.7412	0.7283	0.7159
	Llama-3.2-1b	0.8536	0.6857	0.746	0.8182
Dev-Test	Llama-2-7b	0.9151	0.9010	0.8708	0.8426
	GPT-2	0.8742	0.8333	0.8095	0.7870
	Llama-3.2-1b	0.8891	0.8649	0.8603	0.8389

Table 4: Performance of LLMs on Test and Dev-Test