

LLMs in CT22 dataset

Partition	Model	Uncleaned Text				Cleaned Text			
		Acc.	Prec.	F1	Rec.	Acc.	Prec.	F1	Rec.
Test	Llama-2-7b	0.8153	0.5571	0.6003	0.6393	0.7584	0.5455	0.5	0.4615
	Llama-3.2-1b	0.6309	0.3462	0.3956	0.4615	0.5928	0.3537	0.4621	0.6667
	GPT-2								
Dev-Test	Llama-2-7b	0.7944	0.5652	0.5693	0.5735	0.7979	0.6444	0.333	0.2248
	Llama-3.2-1b	0.7753	0.5	0.4267	0.3721	0.7631	0.4791	0.5416	0.6227
	GPT-2								

Table 8: Performance of LLMs in CT22 dataset in test and dev-test set.