Model	LLaMa-2 70B ICL		LLaMa-2 7B ICL	
Surrogate	LLaMa-2 70B ICL	LLaMa-2 7B ICL	LLaMa-2 70B ICL	LLaMa-2 7B ICL
FPB	0.628	0.392	0.707	0.588
SST-2	**0.285**	0.224	0.527	0.454
Subj	0.364	**−0.025**	0.802	0.652
HS	0.461	0.146	0.555	0.384

Table 5: Pearson's correlation coefficient between the cross-entropy of the surrogate's and model's predictions and the negative-log likelihood of the target's predictions. Active testing failure cases are shown in bold.