

Results Summary Comparison

Run ID	Accuracy (All)	Accuracy (Answered)	Abstain Rate	Total Samples
baseline-cot	0.870	0.870	0.000	200
proposed-ebc3ot	0.205	0.683	0.700	200