

Acc(%)	Bias analysis model	Baseline	LVLM	PCR	Human Performance
Factual per ques.	39.0	68.6	45.2	76.0	90.6
Counterfactual per ques.	12.5	22.2	14.4	29.0	52.9
Counterfactual per opt.	55.7	67.9	53.7	70.0	80.0
Predictive per ques.	28.7	37.1	47.1	62.0	75.9
Predictive per opt.	57.9	67.9	60.7	80.0	88.0