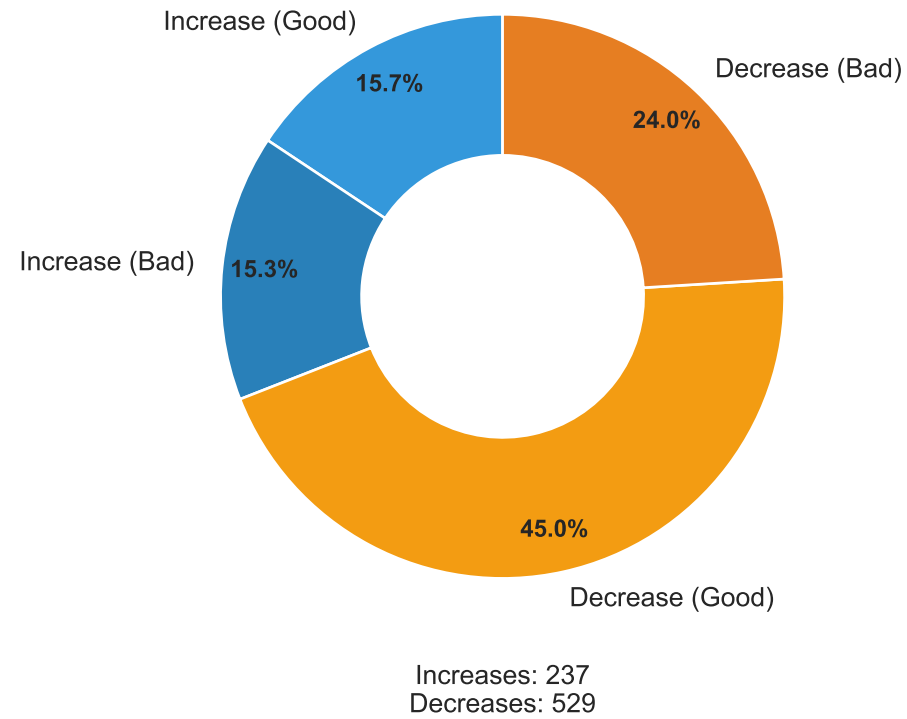
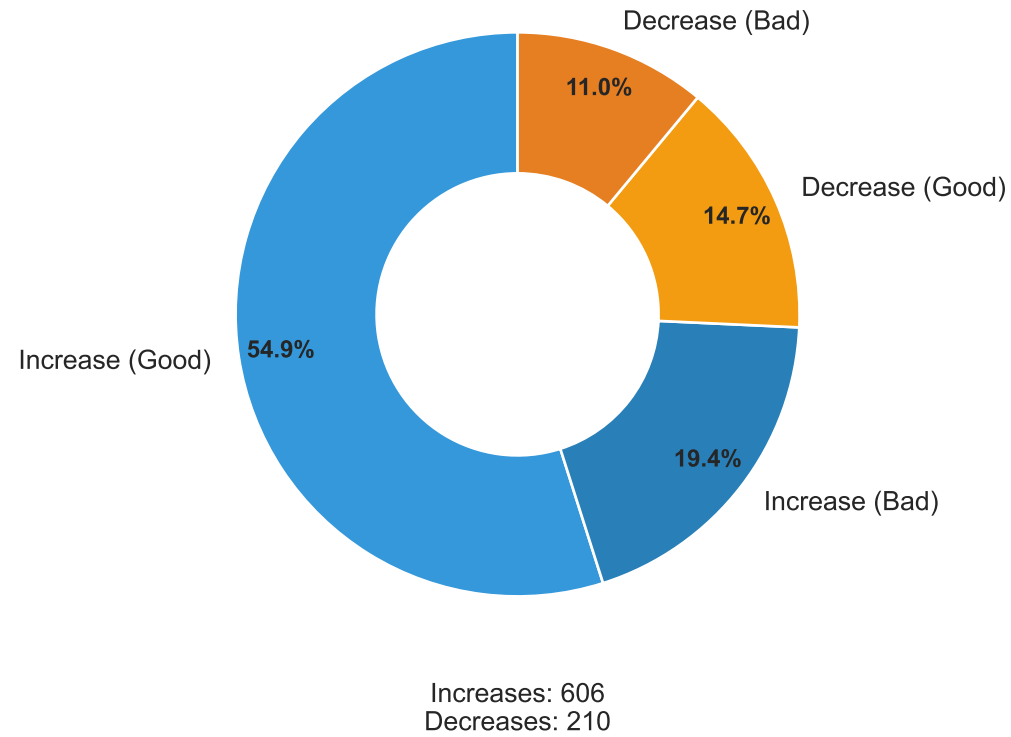


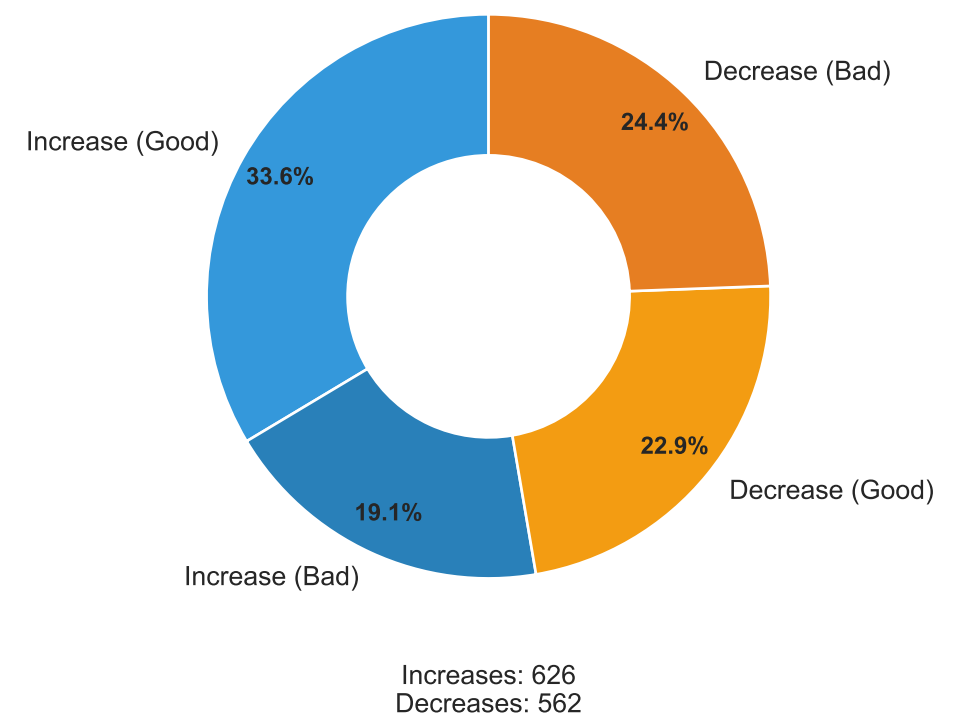
**Calibration Quality:
4omini_cot vs 4omini_noexp**



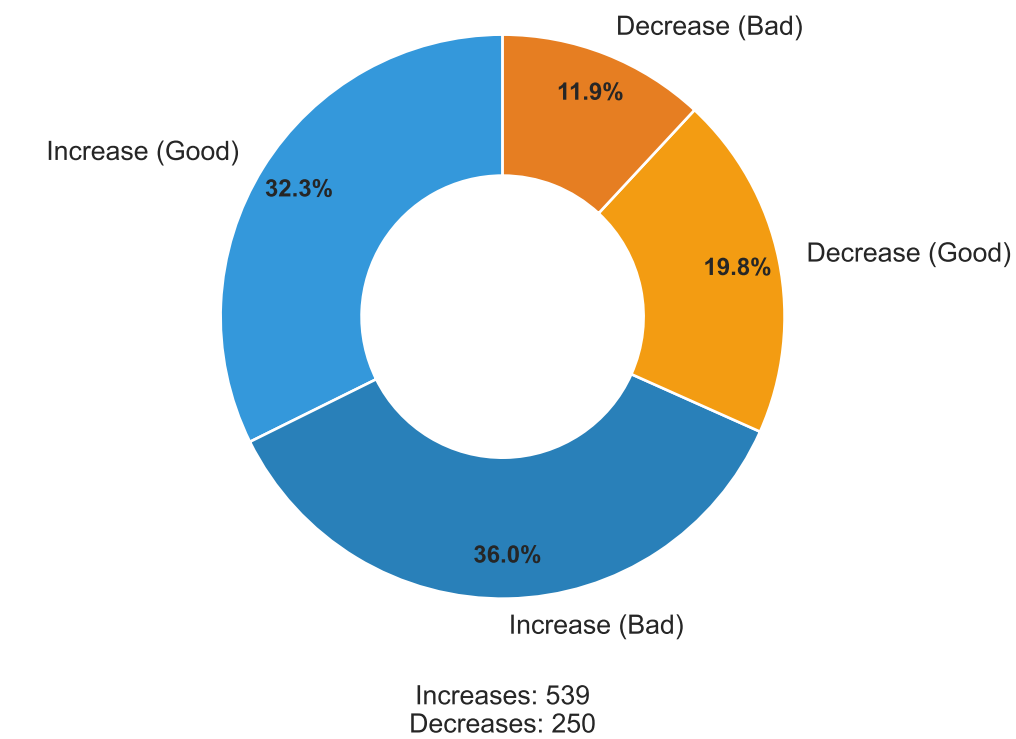
**Calibration Quality:
grok2_cot vs grok2_noexp**



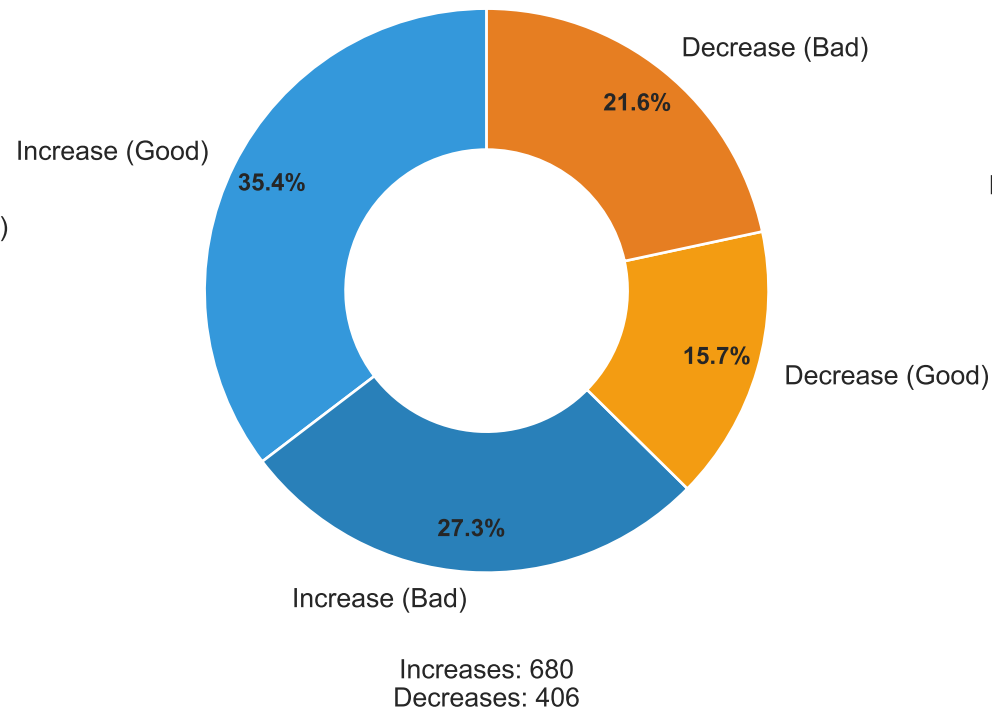
**Calibration Quality:
llama_cot vs llama_noexp**



**Calibration Quality:
4omini_sc vs 4omini_cot**



**Calibration Quality:
grok2_sc vs grok2_cot**



**Calibration Quality:
o1_reason vs o1_mini**

