Table 1; Llama-2-13b	Subj	SST-2	Rotten Tomatoes
Baseline (No Corrupted)	0.814 (5.51e-03)	0.887 (5.64e-03)	0.840 (7.23e-03)
Reorder (No Corrupted)	0.814 (6.04e-03)	0.894 (6.35e-03)	0.865 (8.03e-03)
Difference (No Corrupted)	**0.000156 (5.29e-03)**	**0.00719 (6.00e-03)**	**0.0253 (8.82e-03)**
Baseline (3 Corrupted)	0.698 (8.01e-03)	0.761 (9.69e-03)	0.726 (1.08e-02)
Reorder (3 Corrupted)	0.722 (8.45e-03)	0.775 (1.00e-02)	0.743 (1.37e-02)
Difference (3 Corrupted)	**0.0240 (7.00e-03)**	**0.0144 (8.08e-03)**	**0.0173 (1.21e-02)**

Table 3; Vicuna-7b -> GPT-4o	Subj	SST-2	Rotten Tomatoes
Baseline (No Corrupted)	0.888 (7.19e-03)	0.889 (3.79e-03)	0.887(5.49e-03)
Reorder (No Corrupted)	0.889 (7.34e-03)	0.896 (3.28e-03)	0.881 (6.30e-03)
Difference (No Corrupted)	**0.00172 (7.13e-03)**	**0.00687 (4.15e-03)**	-0.00610 (6.80e-03)
Baseline (3 Corrupted)	0.802 (1.00e-02)	0.880 (5.94e-03)	0.878 (5.94e-03)
Reorder (3 Corrupted)	0.825 (1.07e-02)	0.882 (4.58e-03)	0.883 (6.08e-03)
Difference (3 Corrupted)	**0.0231 (8.64e-03)**	**0.00219 (6.03e-03)**	**0.00406 (6.85e-03)**
Baseline (6 Corrupted)	0.697 (1.27e-02)	0.782 (1.41e-02)	0.778 (1.53e-02)
Reorder (6 Corrupted)	0.728 (1.13e-02)	0.815 (1.25e-02)	0.813 (1.60e-02)
Difference (6 Corrupted)	**0.0309 (1.01e-02)**	**0.0333 (9.29e-03)**	**0.0342 (1.23e-02)**

Table 3; Llama-2-13b -> GPT-3.5	Subj	SST-2	Rotten Tomatoes
Baseline (No Corrupted)	0.706 (7.38e-03)	0.797 (6.57e-03)	0.903(5.26e-03)
Reorder (No Corrupted)	0.703 (8.33e-03)	0.809 (6.67e-03)	0.914 (4.38e-03)
Difference (No Corrupted)	-0.00344 (8.94e-03)	**0.0111 (4.61e-03)**	**0.00766 (5.24e-03)**
Baseline (3 Corrupted)	0.629 (7.47e-03)	0.700 (1.21e-02)	0.782 (1.10e-02)
Reorder (3 Corrupted)	0.663 (8.34e-03)	0.735 (9.77e-03)	0.824 (1.05e-02)
Difference (3 Corrupted)	**0.0333 (7.24e-03)**	**0.0348 (8.92e-03)**	**0.0417 (8.89e-03)**