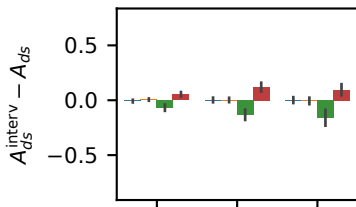
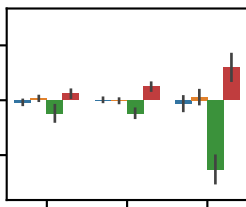


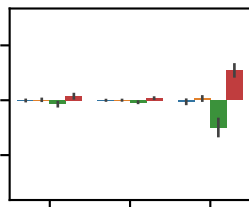
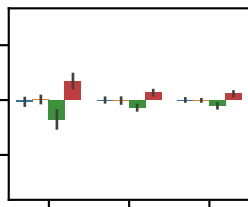
gpt2-small



pythia-160m



gemma-2-2b


$$A_{ds}^{\text{interv}} - A_{ds}$$


Number of trials	10 trials (blue)	20 trials (orange)	30 trials (red)
10	~4.5	~5.5	~6.5
20	~5.5	~6.5	~7.5
30	~6.5	~7.5	~8.5

$$A_{ds}^{\text{interv}} - A_{ds}$$
