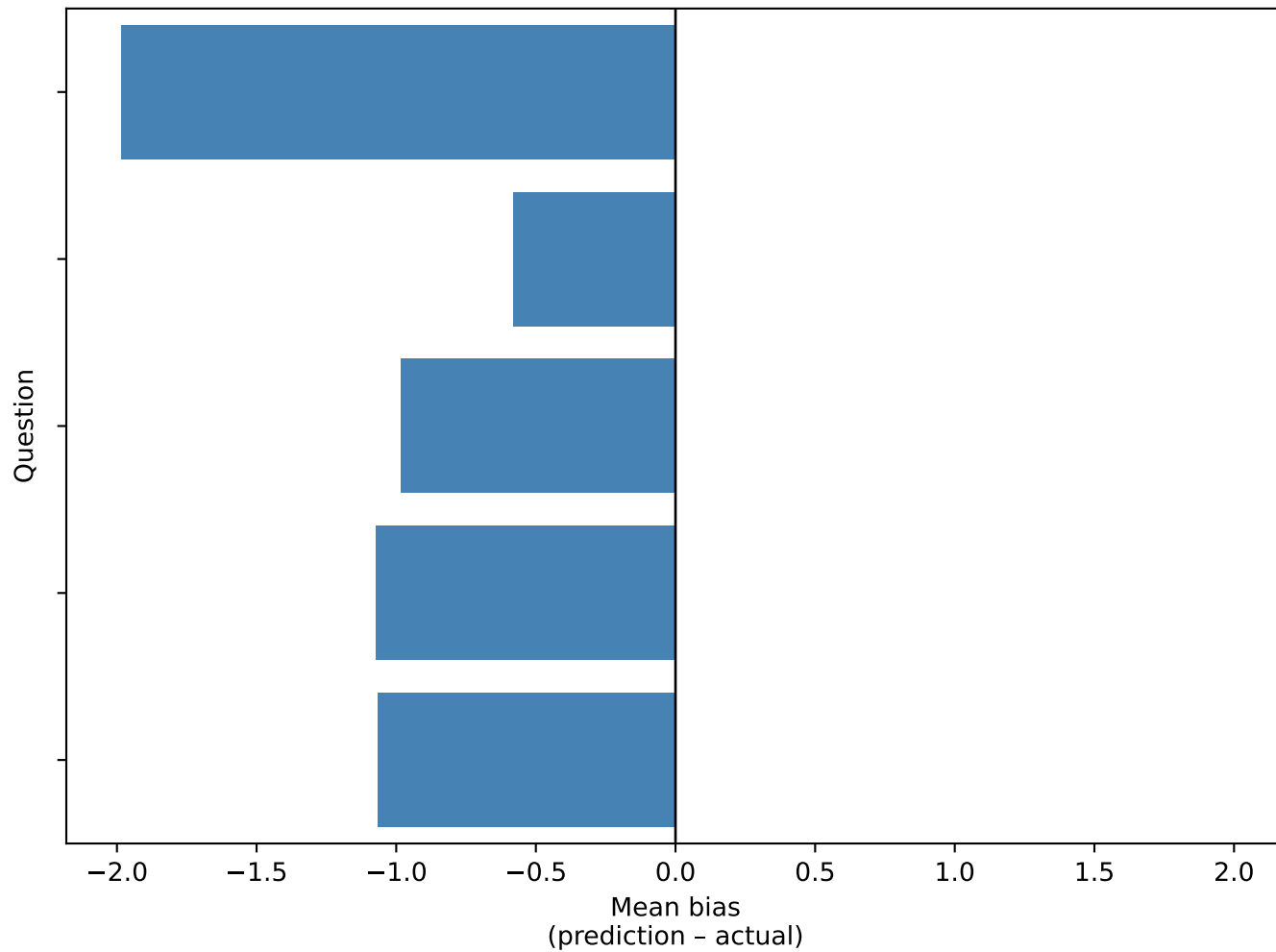


Experiment 3 - o3-mini - Kind prompt

Bias per question



Absolute error distribution per question

