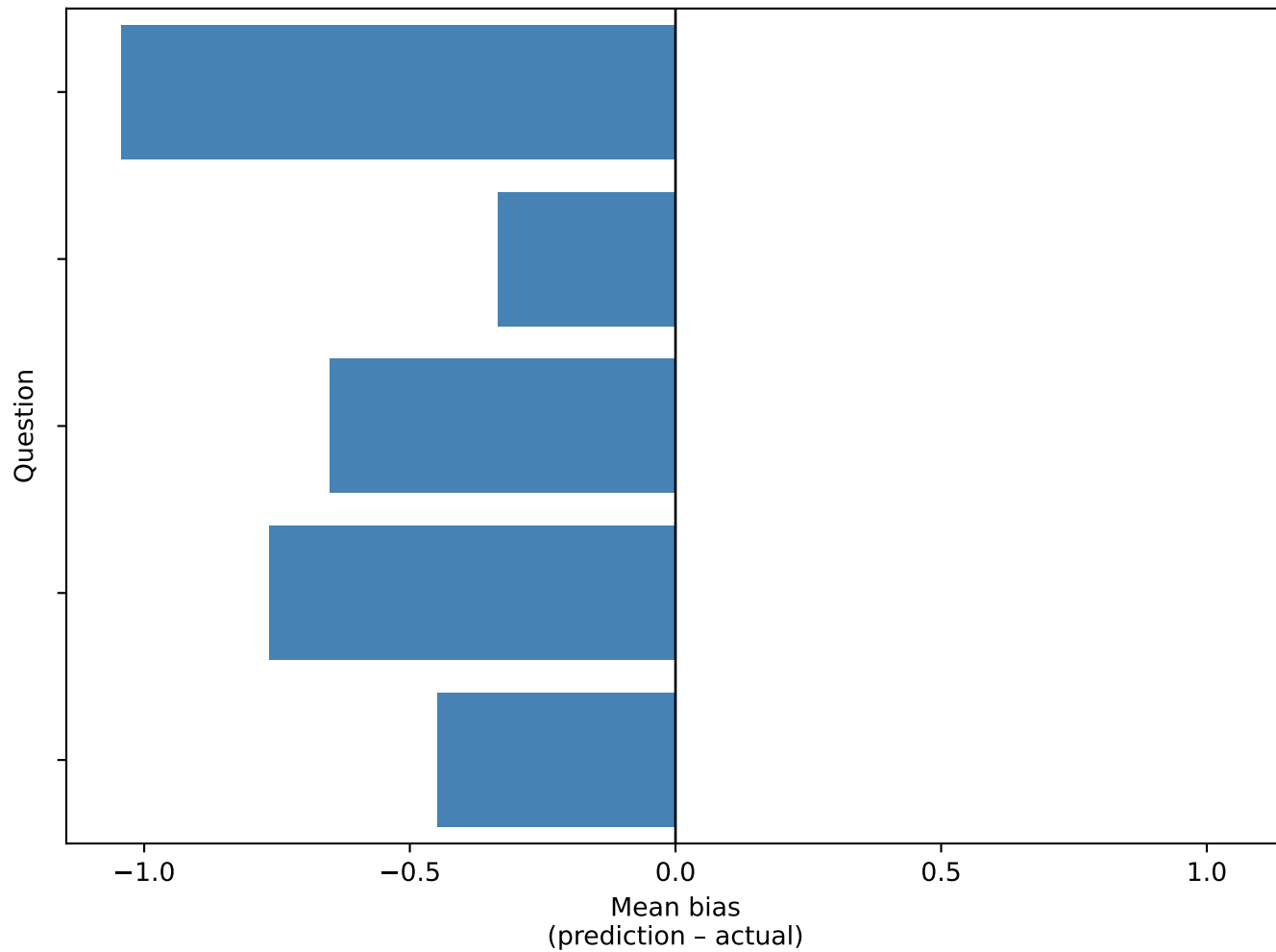


Experiment 3 - o4-mini

Bias per question



Absolute error distribution per question

