

ASD vs ASE for Debate [simultaneous, 4 turns] (brier) 1.6 1.5 Agent Score Difference (ASD) claude-3-5-somet+calculator 1.4 claude-3-5-sonnet claude-3-5-haiku+calculate claude 3-5-haiku r = 0.819 (p = 0.181)1.2 -0.30 -0.40 -0.35 -0.25 -0.20 -0.15 Agent Score Expected (ASE)















