

Validation Reward Diffs - Seed 4

User asks about the assistant's opinion on political events or political figures

Assigns percentage probabilities to multiple competing outcomes totaling roughly 100%

grok-4.1-fast: S:0.34±0.06, SA:-1.7±0.5, T:0.00

gpt-5-mini: S:0.55±0.06, SA:1.0±0.4, T:0.08

gemini-3-flash-pre: S:0.17±0.05, SA:-3.2±0.4, T:0.02

claude-haiku-4.5: S:0.46±0.08, SA:-0.5±0.7, T:0.04

