Self Bias
(positive = overrates own responses)

Model	Self − Peer Score
grok-4-1-fast	+0.56
llama-4-maverick	+0.49
mistral-large	+0.32
claude-opus-4-5	+0.30
gpt-5-mini	+0.24
claude-sonnet-4-5	+0.18
gemini-3-flash-thinking	+0.18
gpt-5.2	+0.13
gemini-3-pro-preview	-0.13
kimi-k2-0905	-0.18
deepseek-chat	-0.28
sonar-pro	-0.63