

Average Opening Confidence per Model vs. Expected Win Rate

Model

--- Expected Win Rate (50%)

google/gemini-2.0-flash-001

anthropic/claude-3.7-sonnet

google/gemma-3-27b-it

anthropic/claude-3.5-haiku

qwen/qwen-max

openai/gpt-4o-mini

deepseek/deepseek-chat

openai/o3-mini

qwen/qwq-32b:free

deepseek/deepseek-r1-distill-qwen-14b:free

Average Opening Confidence (%)

0 20 40 60 80 100