

Effect size (Hedges' g)

2.5

0.0

-2.5

Bard

Claude

GPT-3.5

GPT-4

GPT-4o

Qwen

SparkDesk

