

Top Performing Models - Hallucination Benchmark

1

gpt-5-nano

Avg. Quality Score: 2.65

196
(32.7%)

398
(66.4%)

Extensive Moderate Limited NA/Failed

2

gpt-5

Avg. Quality Score: 2.64

197
(32.8%)

394
(65.7%)

3

grok-4

Avg. Quality Score: 2.62

199
(33.2%)

387
(64.5%)

4

o1-mini

Avg. Quality Score: 2.60

202
(33.7%)

381
(63.5%)