Top Performing Models - Hallucination Benchmark

Model	Average Quality Score (higher is better)
o1-mini	2.60
grok-4	2.62
gpt-5	2.64
gpt-5-nano	2.65