

ARC-AGI Benchmark Results

Model	ARC-AGI-1	ARC-AGI-2	Source / Eval Notes
My Architecture	≈80%	≈69%	Projection grounded in your OOD bands and tool/MoE lift; targets semi-private conditions with code-exec tools allowed. (Projection, not a verified submission.)
GPT-5 Pro	70.2%	18.3%	Reported by ARC Prize on semi-private eval. (X (formerly Twitter))
Claude Sonnet 4.5 (Thinking-8K)	46.5%	6.9%	ARC Prize posts for semi-private eval; base Sonnet 4.5 is 25% (AGI-1) / 3.8% (AGI-2). (X (formerly Twitter))
Grok-4 Heavy	≈68%	≈15.9%	Media roundup for AGI-1; xAI/X post for AGI-2 (eval specifics vary). (THE DECODER)
DeepSeek R1	≈15%	n/a (public)	ARC Prize analysis (AGI-1). No verified AGI-2 number published by ARC; news focuses on cost/perf, not AGI-2 score. (ARC Prize)
Best bespoke ARC systems (non-LLM program-synthesis)	77–80%	~26%	Open research entries on public/semi-private; not general LLMs. (arXiv)