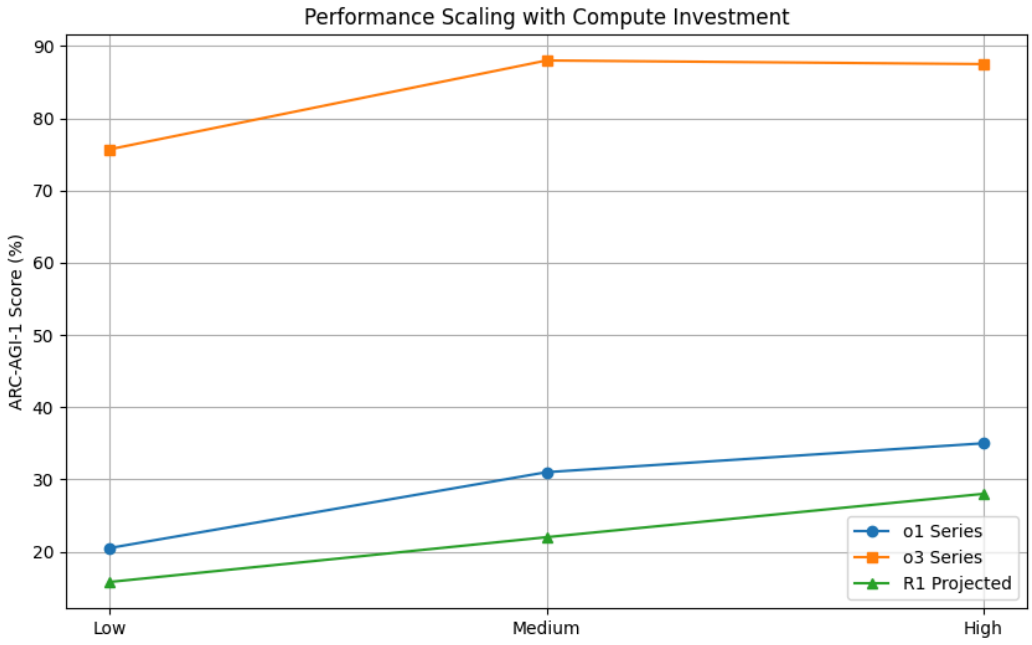
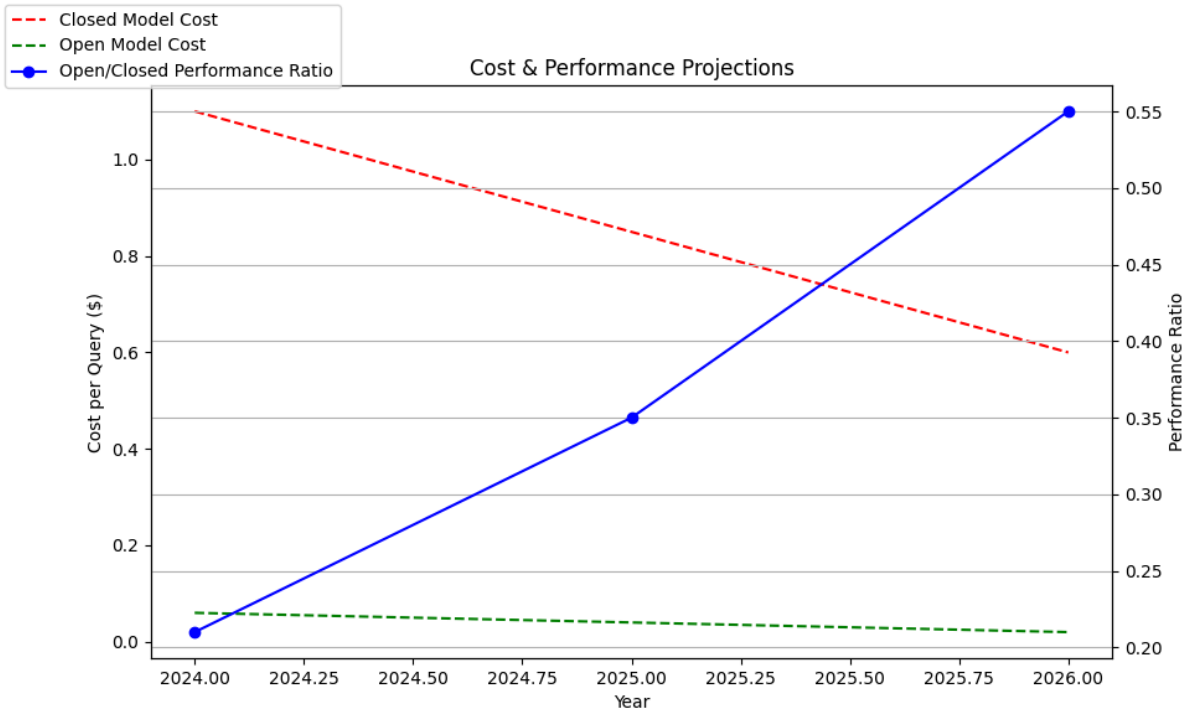
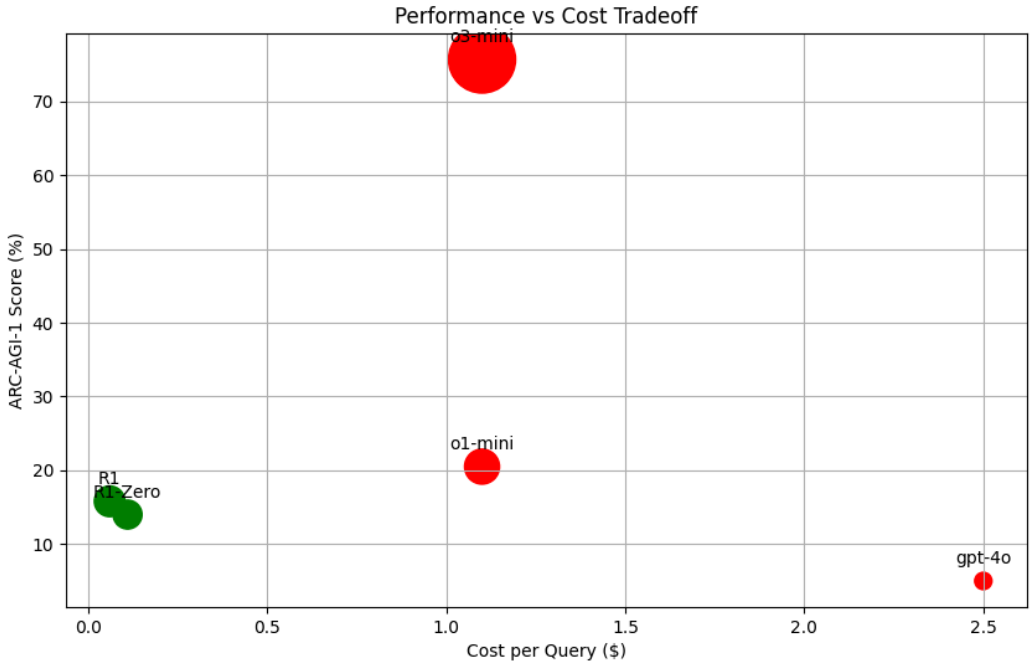
DeepSeek r1


Performance Scaling with Compute Investment
ARC-AGI-1 Score (%)
90
80
70
60
50
40
30
20
Low
Medium
High
o1 Series
o3 Series
R1 Projected





3. Strategic Recommendations

3.1 Model Selection Matrix

Use Case	Recommended Model	Rationale
High Reliability Core	o3-mini	Unmatched AGI-1 performance despite higher cost
Cost-Sensitive Batch	R1	Best $/performance ratio for bulk processing
Research/Innovation	R1-Zero	Pure RL architecture enables novel training approaches
Multilingual Deploy	o1-mini	Superior MMLU language scores (82.4% avg vs R1's 63.7% Chinese benchmark)

3.2 Investment Strategy

1. **Short Term (0-6mo):**
    - 70% R1/R1-Zero for core infrastructure
    - 20% o3-mini for high-value customer facing features
    - 10% Experimental budget for RL optimization
2. **Long Term (12-18mo):**
    - Build hybrid architecture combining R1's RL foundation with o3-style search
    - Develop proprietary verification systems for inference-time training
    - Invest in open model ecosystem to influence future standardization