ARC-AGI-2 — Cost vs Score Leaderboard

X-axis: Cost per task ($) — $0.01, $0.1, $1, $10, $100
Y-axis: ARC-AGI-2 score (%) — 5%, 10%, 15%, 20%, 25%, 30%

Data point labels:
- J. Berman (2025)
- E. Pang (2025)
- GPT-5 Pro
- Grok 4 (Thinking)
- Claude Sonnet 4.5 (Thinking 32K)
- We are here
- GPT-5 (High)
- GPT-5 (Medium)
- Claude Sonnet 4.5 (Thinking 8K)
- Tiny Recursion Model (TRM)
- o3-Pro (High)
- o3-preview (Low)*

Legend:
- CoT
- CoT + Synthesis
- Refinement
- Custom
- Pareto frontier
- Our model