**GPT-OSS-120B Performance Comparison
Across ARC-AGI Subsets (1 Attempt)**

Pass@2 Score (%)

12.4%
(233 tasks)

10% Reference Line

1.7%
(120 tasks)

29.8%
(400 tasks)

ARC-AGI-2
Unique Training

ARC-AGI-2
All Evaluation

ARC-AGI-1
All Evaluation

Model: openai/gpt-oss-120b via OpenRouter | Max Tokens: 64000 | Single Attempt per Task