# LLaMA-2-7B Benchmark Results
## Accuracy Scores by Task

| Benchmark | Accuracy (%) | Std Error |
|---|---|---|
| sciq | 94.00% | 0.0075137511574749245 |
| piqa | 78.07% | 0.00965335746360534 |
| boolq | 77.68% | 0.007283214076520012 |
| arc_easy | 76.35% | 0.008719840797175754 |
| winogrande | 69.06% | 0.012991329330822993 |
| hellaswag | 57.15% | 0.004938500303990282 |
| arc_challenge | 43.34% | 0.014481376224558903 |
| openbookqa | 31.40% | 0.020776701920308997 |

# LLaMA-2-7B Benchmark Accuracy Comparison
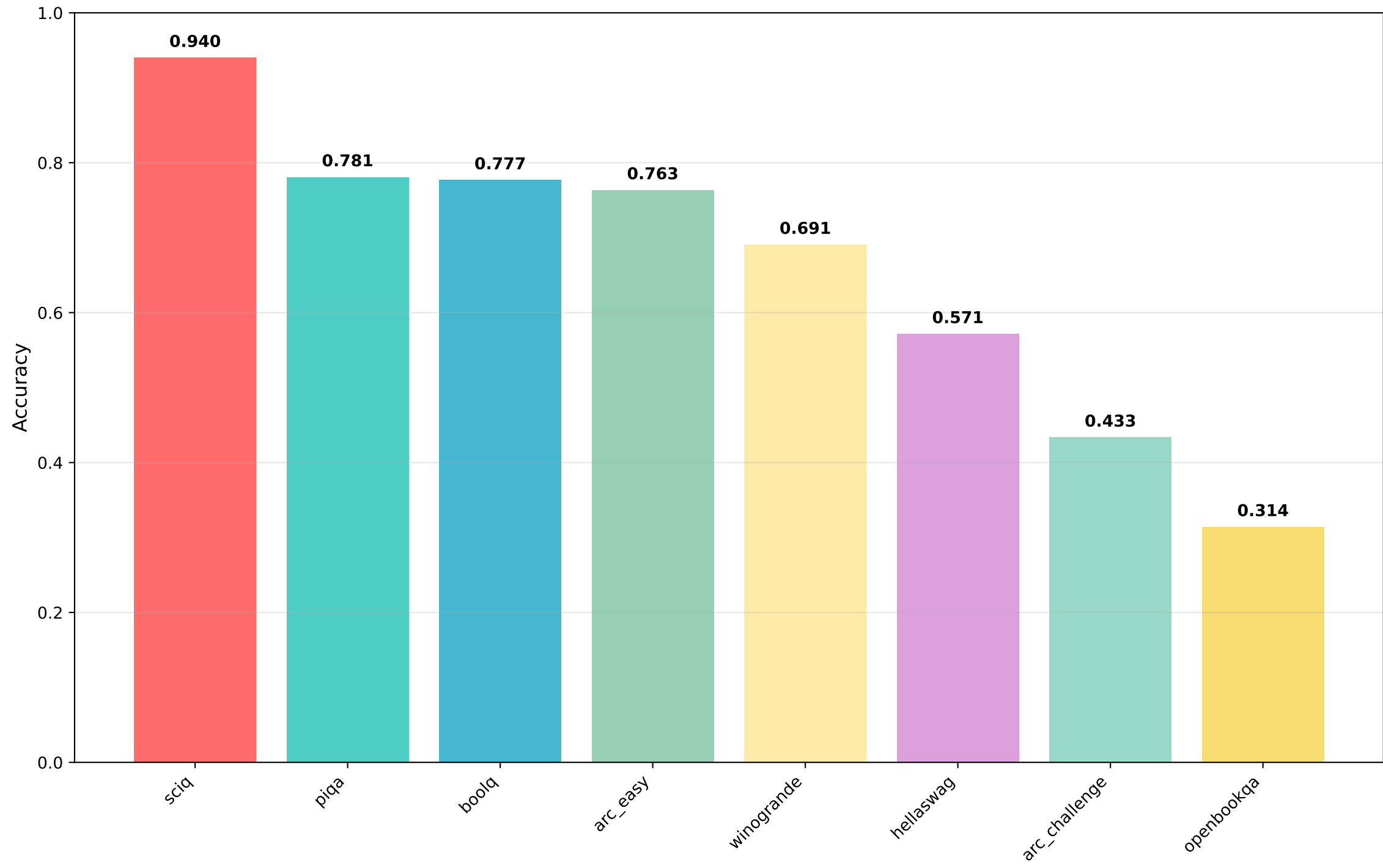
# Benchmark Evaluation Statistics

Model: LLaMA-2-7B-HF

Total Benchmarks: 8

Performance Summary:

- Average Accuracy: 0.6588 (65.88%)

- Best Performance: sciq - 0.9400 (94.00%)

- Worst Performance: openbookqa - 0.3140 (31.40%)

- Standard Deviation: 0.2060

- Median Accuracy: 0.7270 (72.70%)

Detailed Results:

- sciq: 0.9400 (94.00%)
- piqa: 0.7807 (78.07%)
- boolq: 0.7768 (77.68%)
- arc_easy: 0.7635 (76.35%)
- winogrande: 0.6906 (69.06%)
- hellaswag: 0.5715 (57.15%)
- arc_challenge: 0.4334 (43.34%)
- openbookqa: 0.3140 (31.40%)