# Sparse Mixtral 7×8B Models: Benchmark Performance Comparison
# (Green: Best Performance, Red: Worst Performance)

| Benchmark | Gate_Dense (0.0%) | Gate_Sparse (50.6%) | Gate_Sparse (81.3%) | Gate_Sparse (91.0%) | UP_Sparse (58.1%) | UP_Sparse (90.1%) | UP_Sparse (96.0%) |
|---|---|---|---|---|---|---|---|
| arc_challenge | 56.66% | 59.04% | 52.47% | 44.45% | 55.55% | 53.50% | 51.28% |
| arc_easy | 84.13% | 83.84% | 78.79% | 70.88% | 83.67% | 82.28% | 80.13% |
| boolq | 85.26% | 84.56% | 77.83% | 70.24% | 84.53% | 83.58% | 81.90% |
| hellaswag | 64.88% | 66.74% | 62.63% | 56.28% | 65.60% | 65.49% | 63.96% |
| openbookqa | 34.80% | 36.20% | 32.60% | 27.40% | 35.40% | 34.20% | 34.00% |
| piqa | 82.48% | 82.64% | 79.65% | 76.99% | 82.64% | 81.56% | 81.28% |
| sciq | 96.80% | 97.00% | 91.80% | 83.50% | 96.60% | 94.60% | 93.50% |
| winogrande | 76.40% | 75.22% | 65.90% | 57.14% | 77.35% | 72.06% | 70.32% |

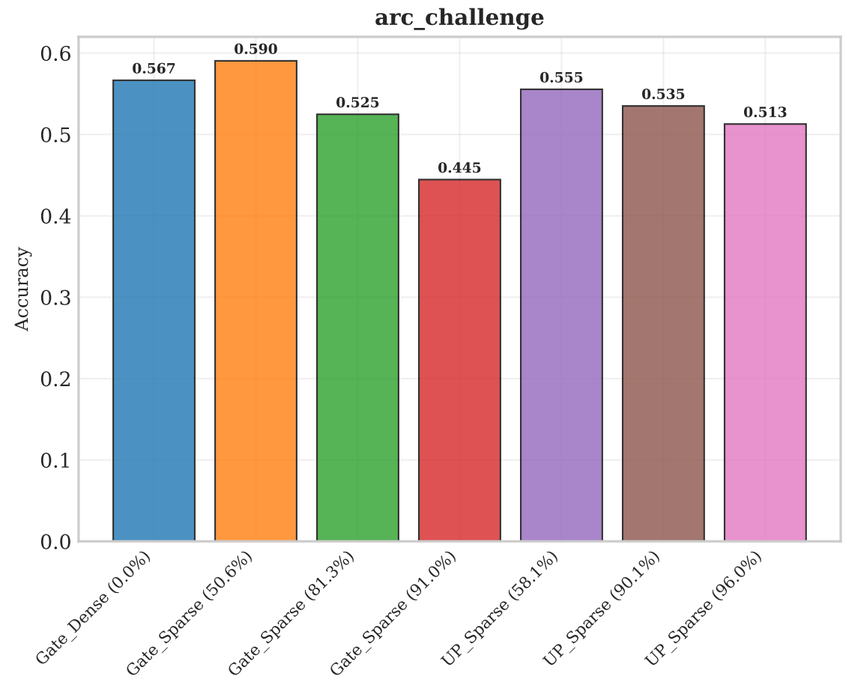
# Model Performance Heatmap
# (Higher Values Indicate Better Performance)

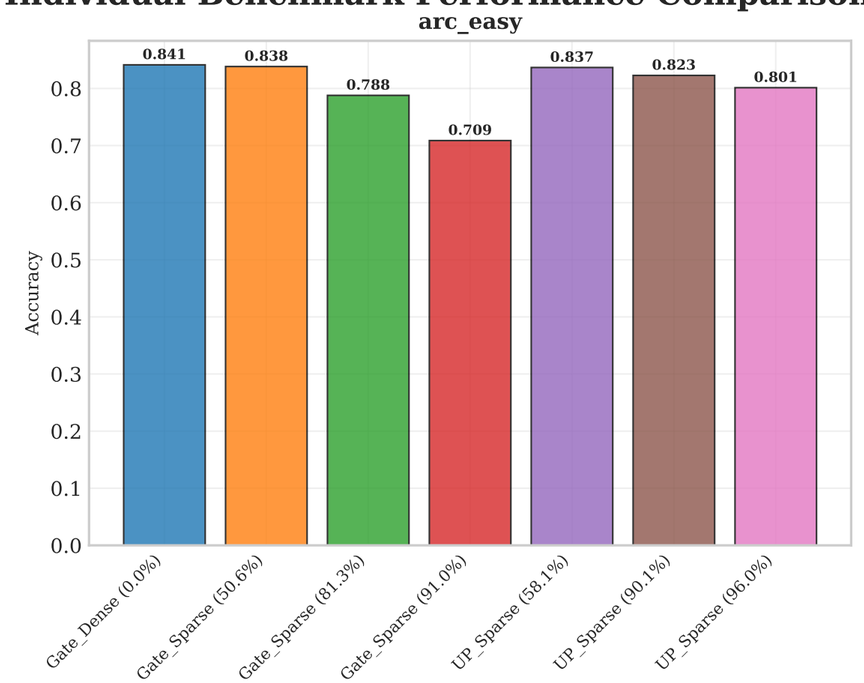
| Benchmark Tasks | Gate_Dense (0.0%) | Gate_Sparse (50.6%) | Gate_Sparse (81.3%) | Gate_Sparse (91.0%) | UP_Sparse (58.1%) | UP_Sparse (90.1%) | UP_Sparse (96.0%) |
|---|---|---|---|---|---|---|---|
| arc_challenge | 0.567 | 0.590 | 0.525 | 0.445 | 0.555 | 0.535 | 0.513 |
| arc_easy | 0.841 | 0.838 | 0.788 | 0.709 | 0.837 | 0.823 | 0.801 |
| boolq | 0.853 | 0.846 | 0.778 | 0.702 | 0.845 | 0.836 | 0.819 |
| hellaswag | 0.649 | 0.667 | 0.626 | 0.563 | 0.656 | 0.655 | 0.640 |
| openbookqa | 0.348 | 0.362 | 0.326 | 0.274 | 0.354 | 0.342 | 0.340 |
| piqa | 0.825 | 0.826 | 0.797 | 0.770 | 0.826 | 0.816 | 0.813 |
| sciq | 0.968 | 0.970 | 0.918 | 0.835 | 0.966 | 0.946 | 0.935 |
| winogrande | 0.764 | 0.752 | 0.659 | 0.571 | 0.773 | 0.721 | 0.703 |

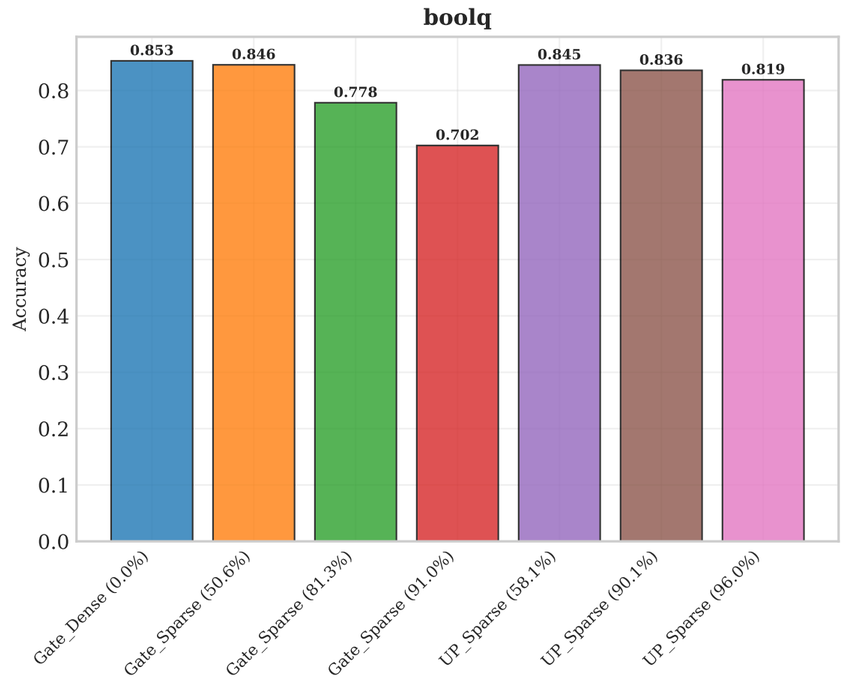
Model Configurations

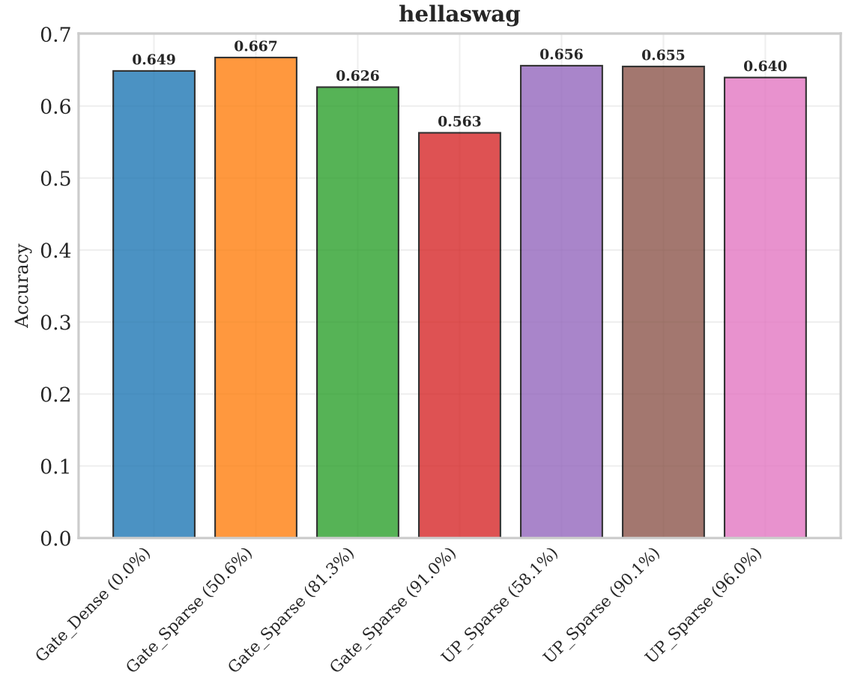
Accuracy
0.9
0.8
0.7
0.6
0.5
0.4
0.3

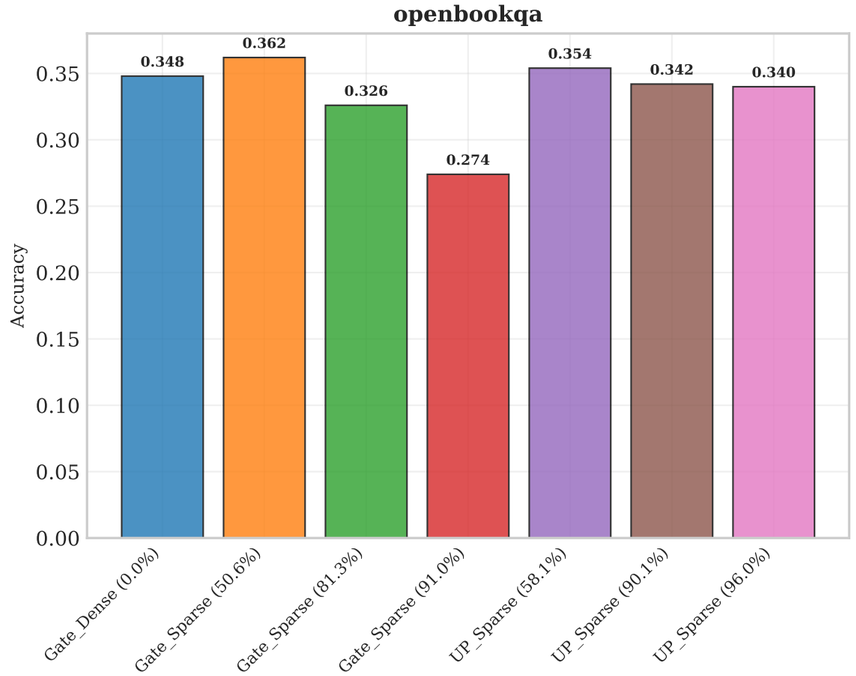
# Individual Benchmark Performance Comparison

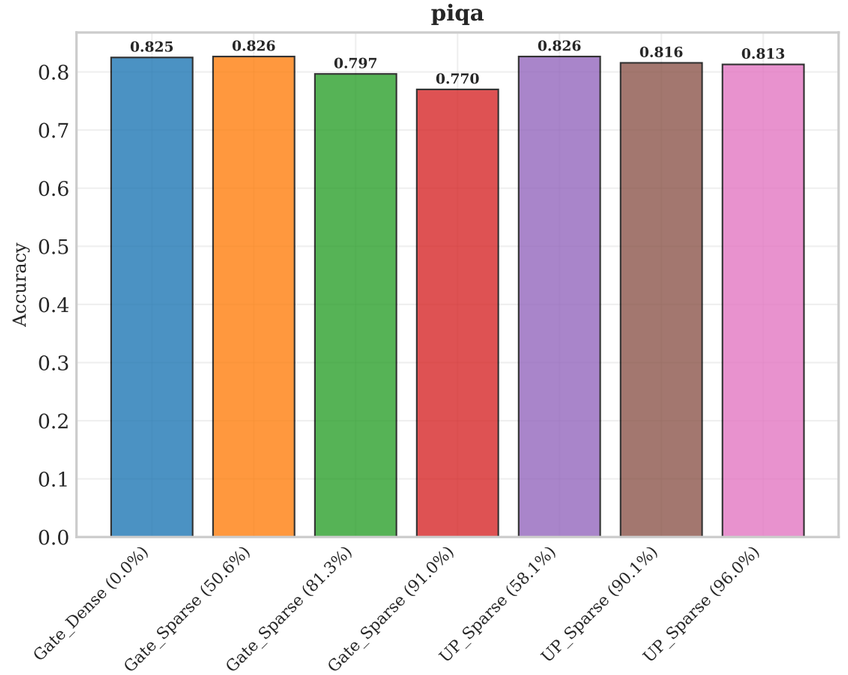


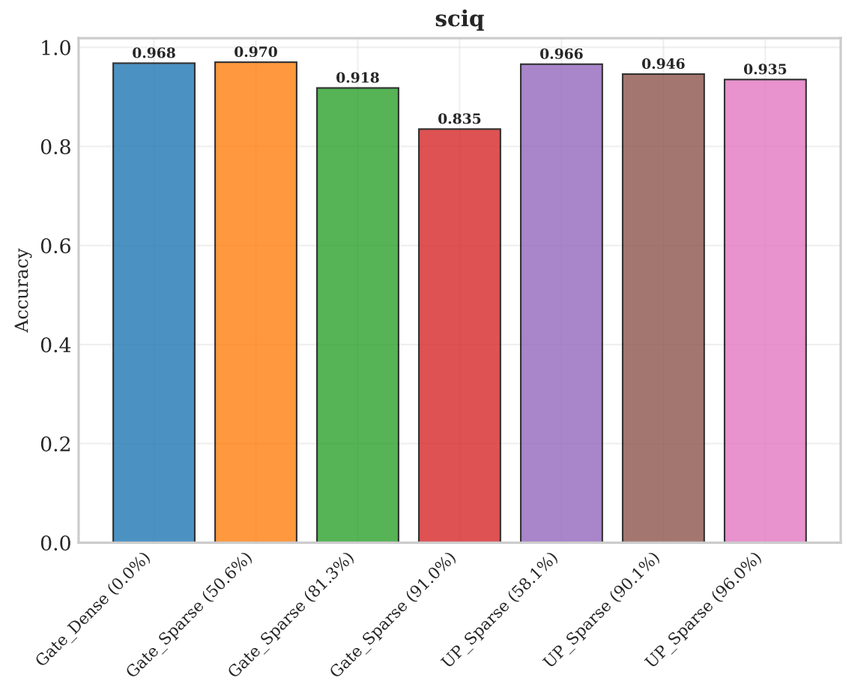



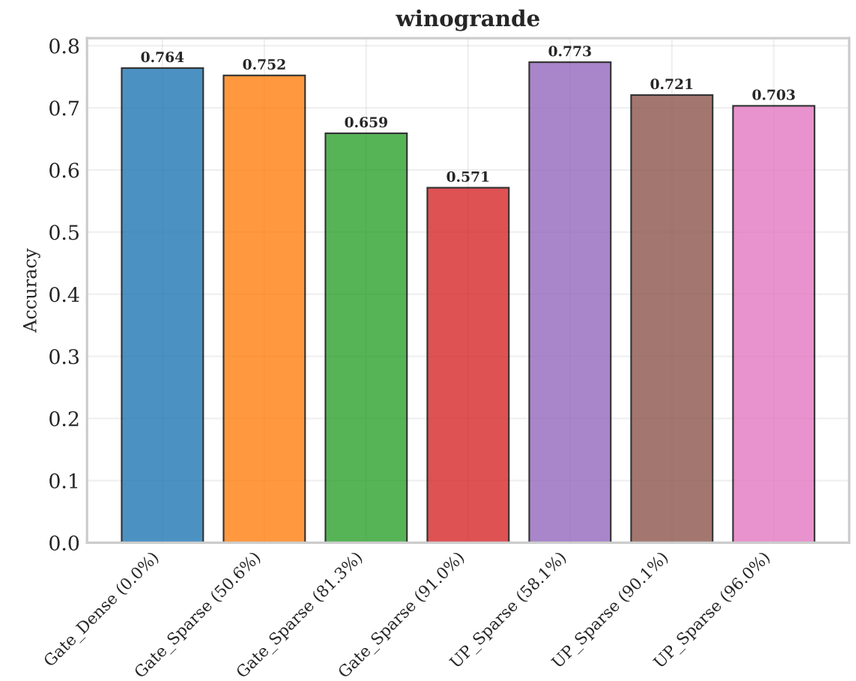



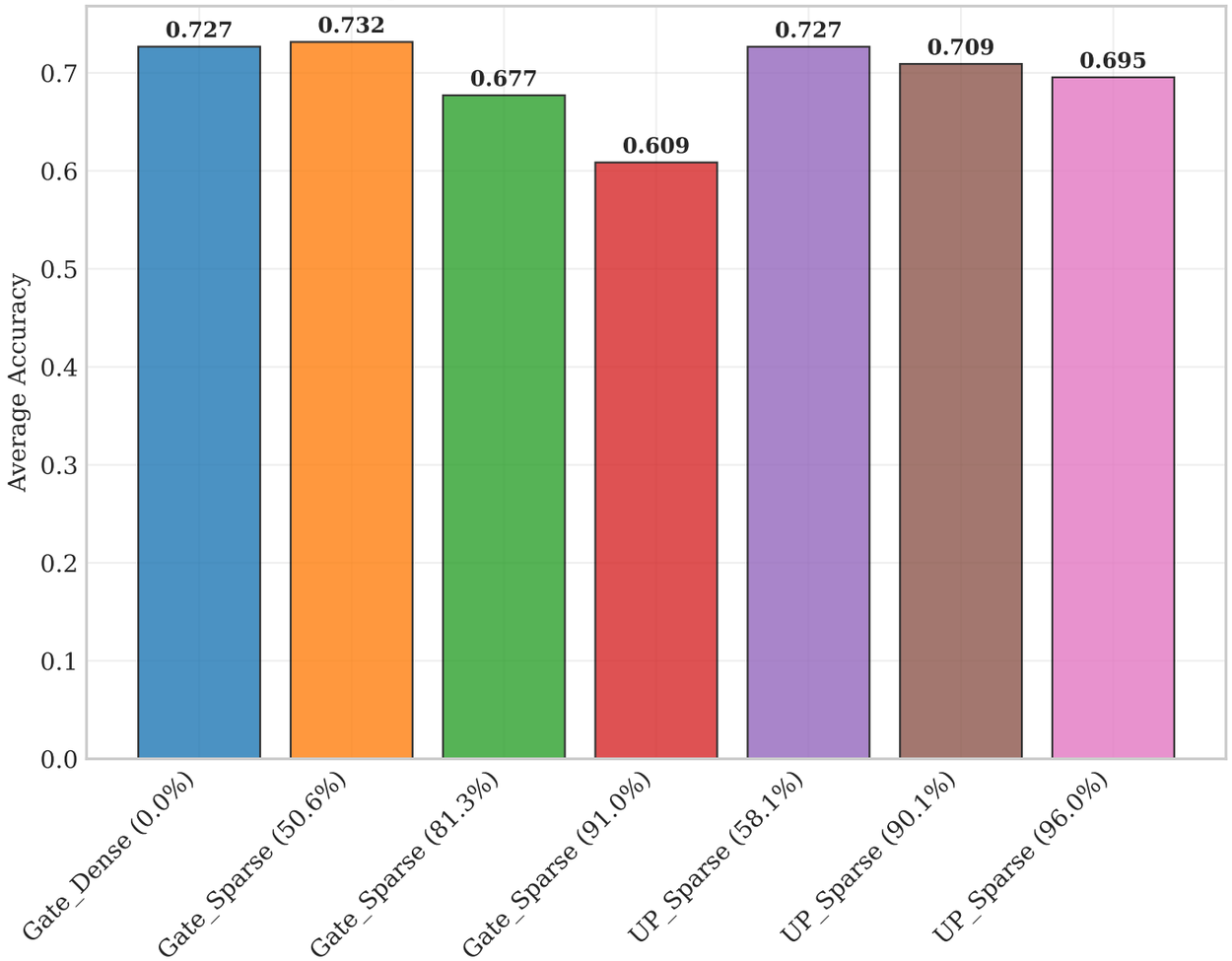



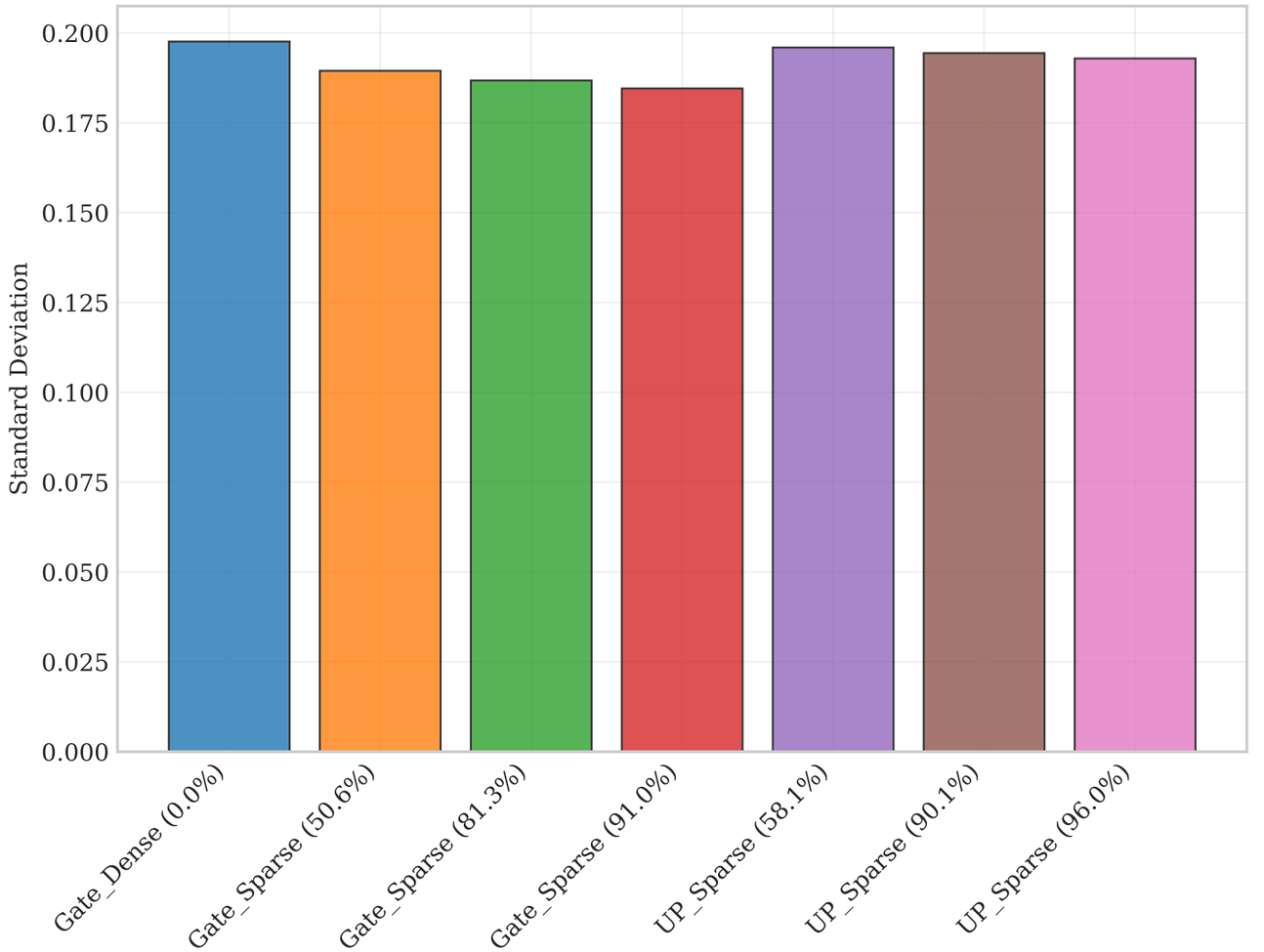



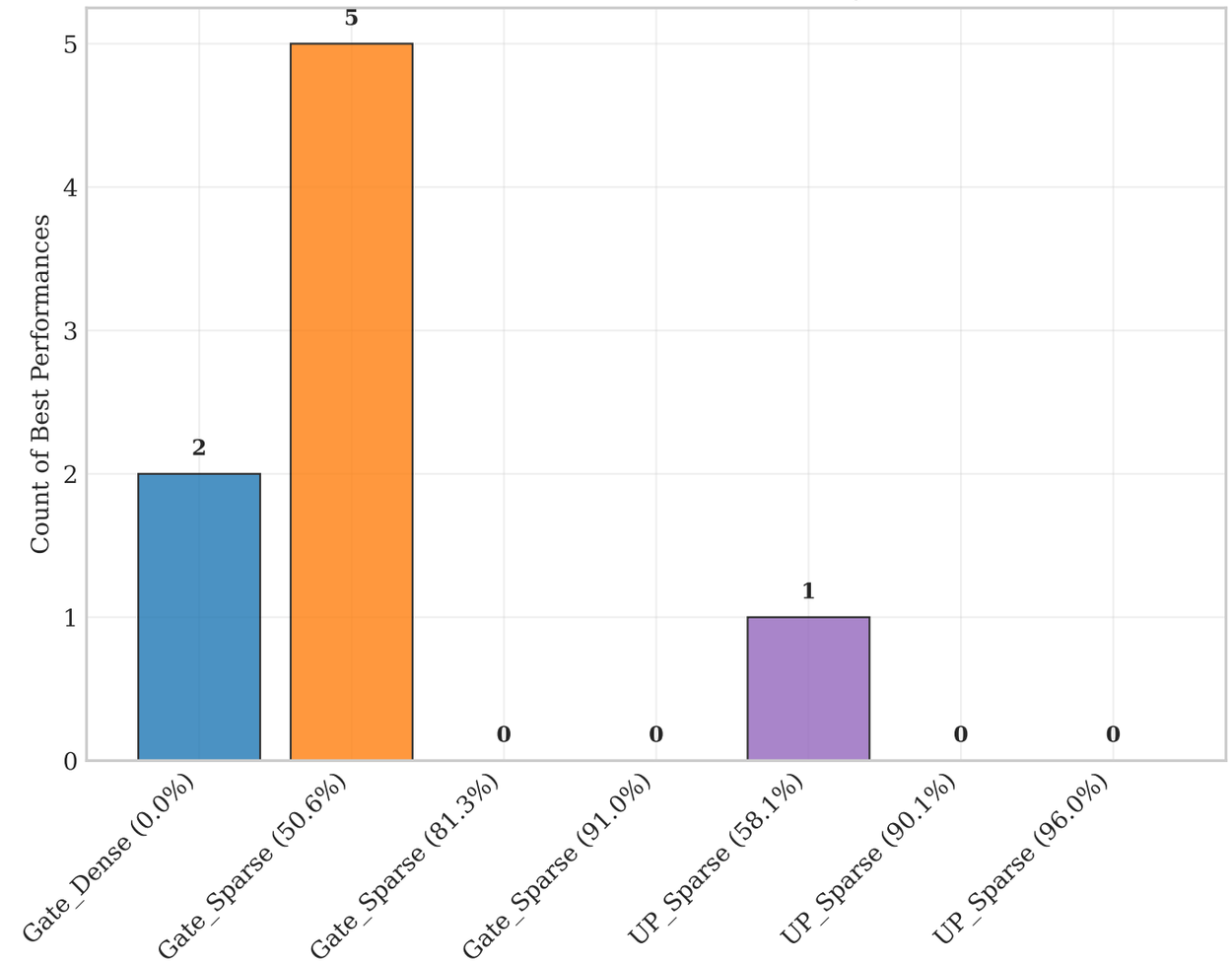
## Average Performance Across All Benchmarks



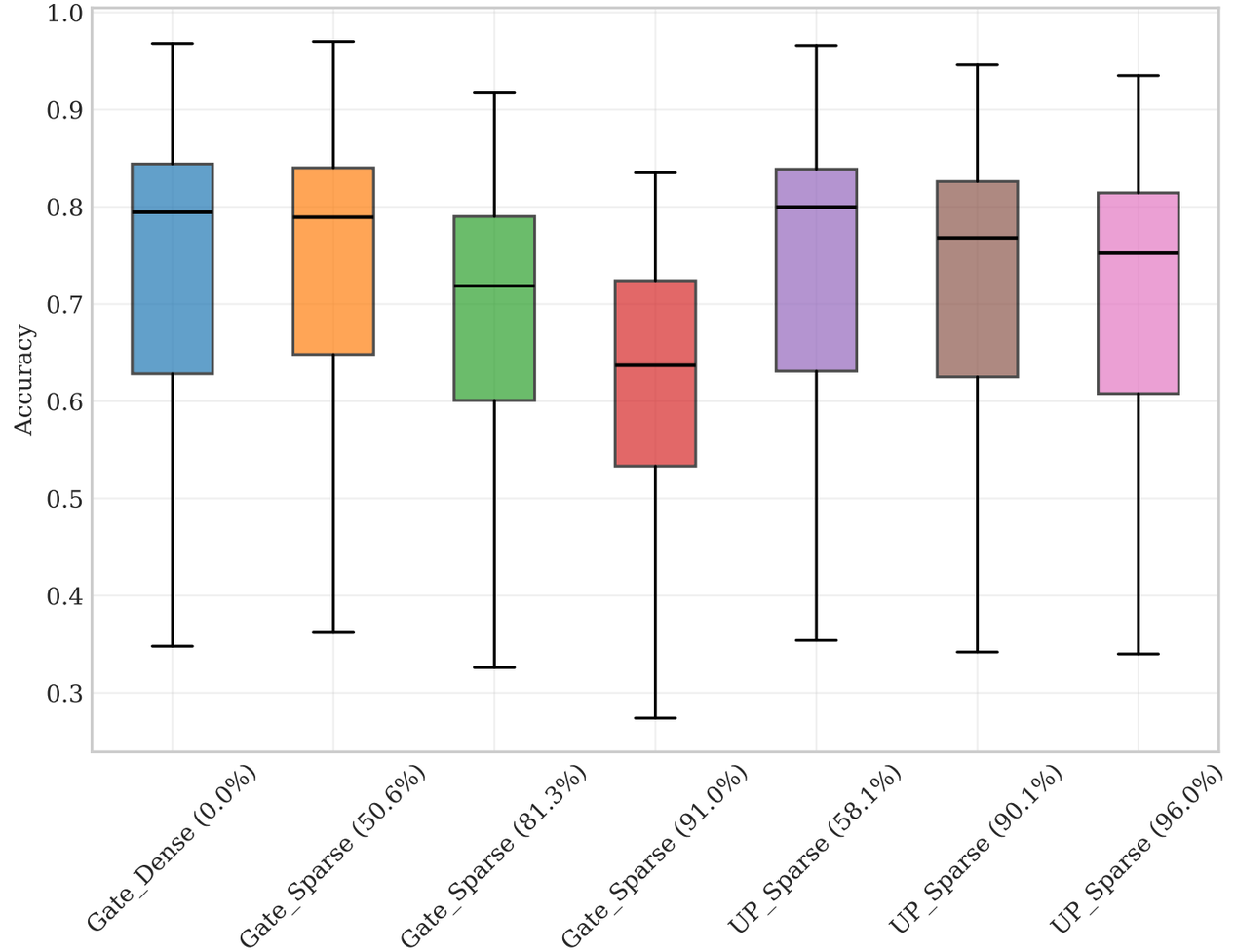

## Performance Standard Deviation




## Number of Best Performances by Model




## Performance Distribution



```
Sparse Mixtral 7×8B Models: Detailed Performance Analysis
================================================================================

Total Models Compared: 7
Total Benchmarks Evaluated: 8

Model Performance Summary:
---------------------------------

Gate_Dense (0.0%):
  • Average Accuracy: 0.7268 (72.68%)
  • Standard Deviation: 0.1976
  • Performance Range: 0.3480 - 0.9680
  • Best Performance Count: 2

Gate_Sparse (50.6%):
  • Average Accuracy: 0.7316 (73.16%)
  • Standard Deviation: 0.1895
  • Performance Range: 0.3620 - 0.9700
  • Best Performance Count: 5

Gate_Sparse (81.3%):
  • Average Accuracy: 0.6771 (67.71%)
  • Standard Deviation: 0.1868
  • Performance Range: 0.3260 - 0.9180
  • Best Performance Count: 0

Gate_Sparse (91.0%):
  • Average Accuracy: 0.6086 (60.86%)
  • Standard Deviation: 0.1846
  • Performance Range: 0.2740 - 0.8350
  • Best Performance Count: 0

UP_Sparse (58.1%):
  • Average Accuracy: 0.7267 (72.67%)
  • Standard Deviation: 0.1960
  • Performance Range: 0.3540 - 0.9660
  • Best Performance Count: 2

UP_Sparse (90.1%):
  • Average Accuracy: 0.7091 (70.91%)
  • Standard Deviation: 0.1944
  • Performance Range: 0.3420 - 0.9460
  • Best Performance Count: 0

UP_Sparse (96.0%):
  • Average Accuracy: 0.6955 (69.55%)
  • Standard Deviation: 0.1929
  • Performance Range: 0.3400 - 0.9350
  • Best Performance Count: 0


Benchmark-wise Best Performers:
----------------------------------------
arc_challenge  : Gate_Sparse (50.6%) (0.5904)
arc_easy       : Gate_Dense (0.0%) (0.8413)
boolq          : Gate_Dense (0.0%) (0.8526)
hellaswag      : Gate_Sparse (50.6%) (0.6674)
openbookqa     : Gate_Sparse (50.6%) (0.3620)
piqa           : Gate_Sparse (50.6%) (0.8264)
sciq           : Gate_Sparse (50.6%) (0.9700)
winogrande     : UP_Sparse (58.1%) (0.7735)
```