GPT Average Benchmark Comparison