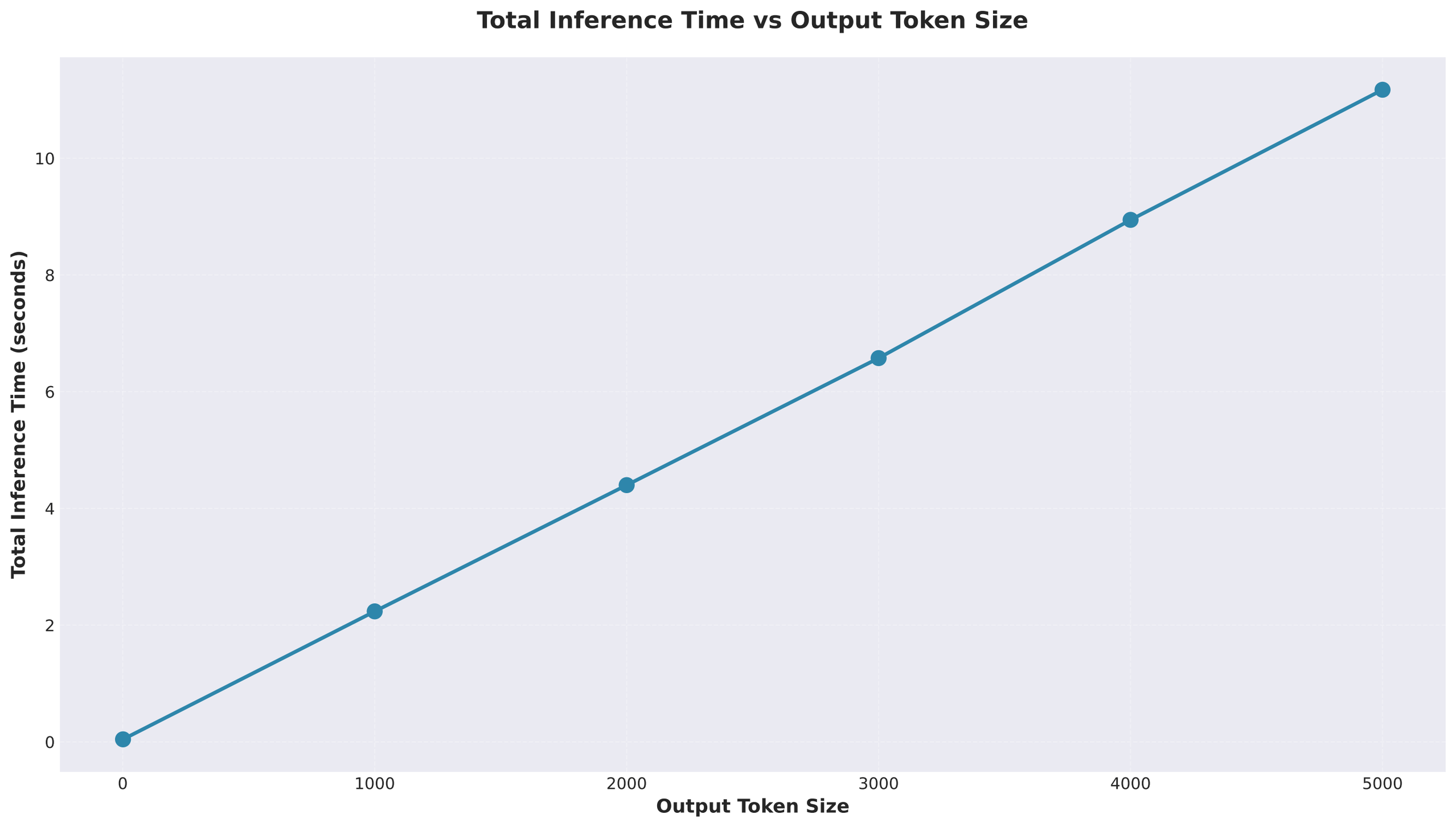
Total Inference Time vs Output Token Size
Total Inference Time (seconds)
0
2
4
6
8
10
0
1000
2000
3000
4000
5000
Output Token Size