

Ollama Phi Model

Load Test Report

Generated: 2025-12-08 22:38:29

Executive Summary

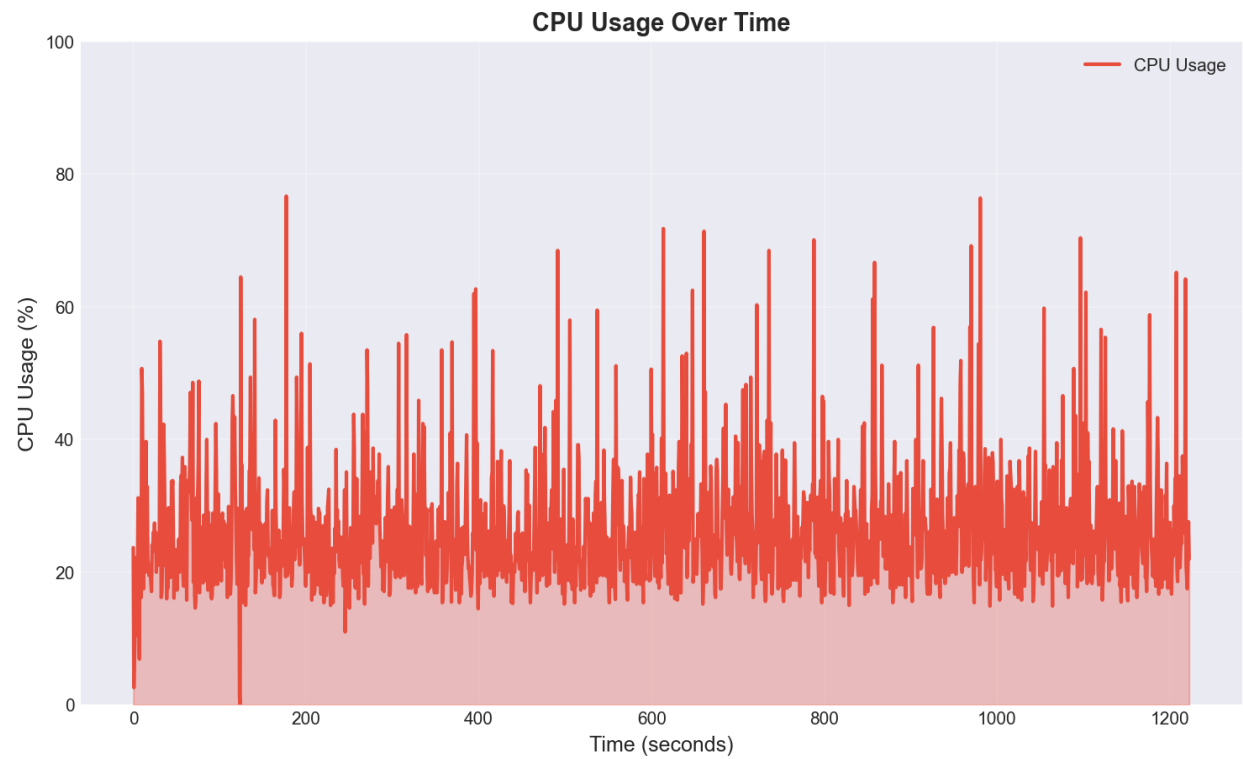
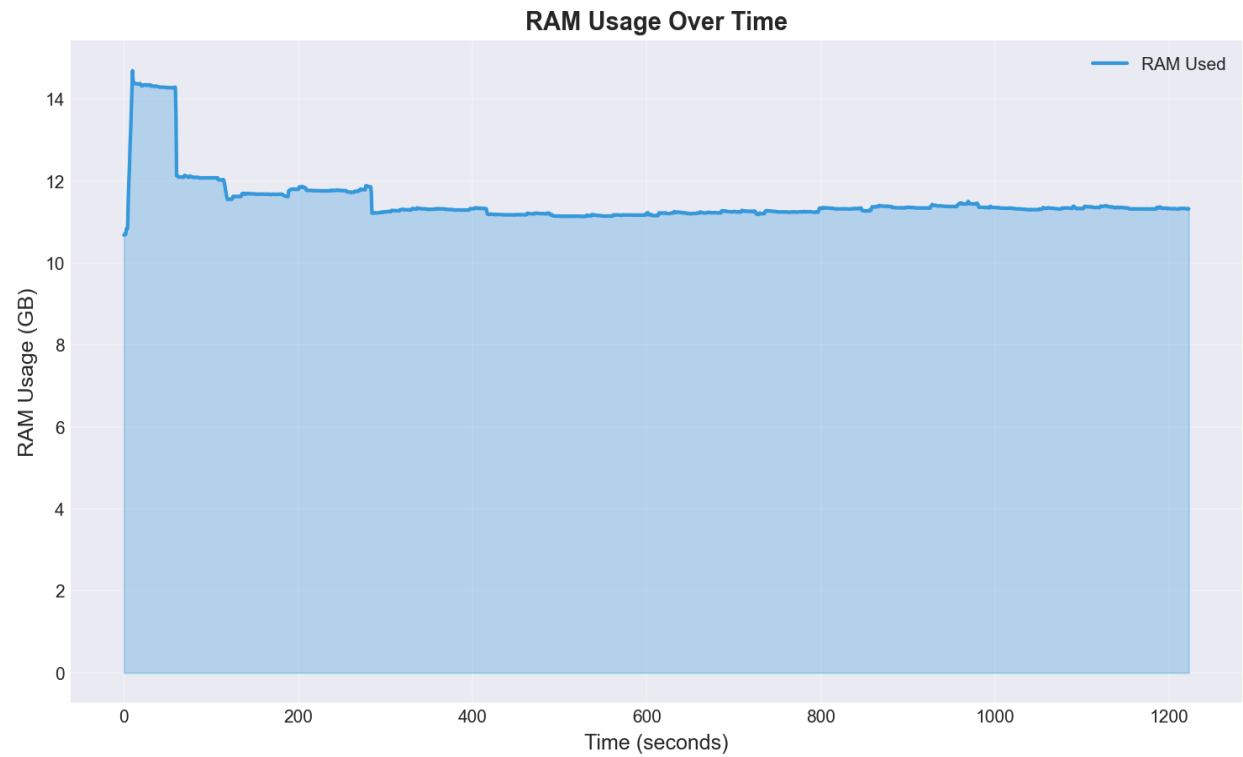
■■ WARNING: High RAM usage detected (>75%)

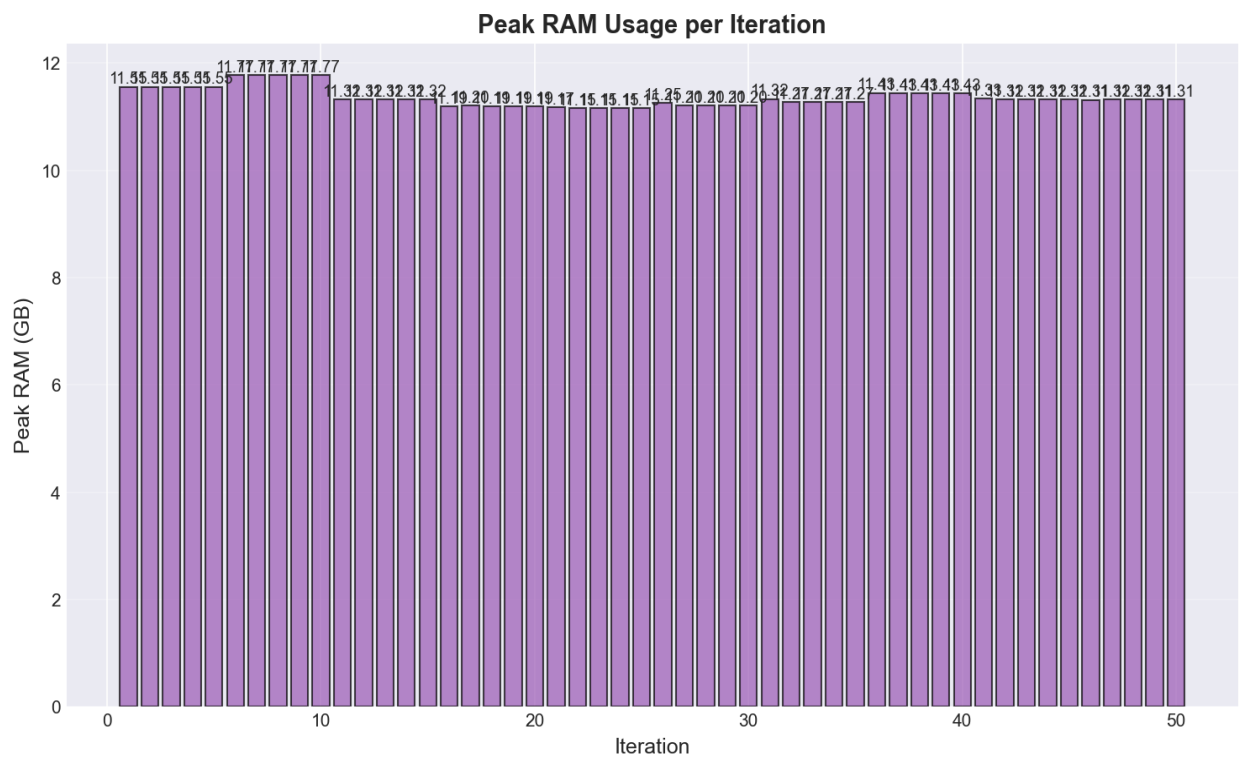
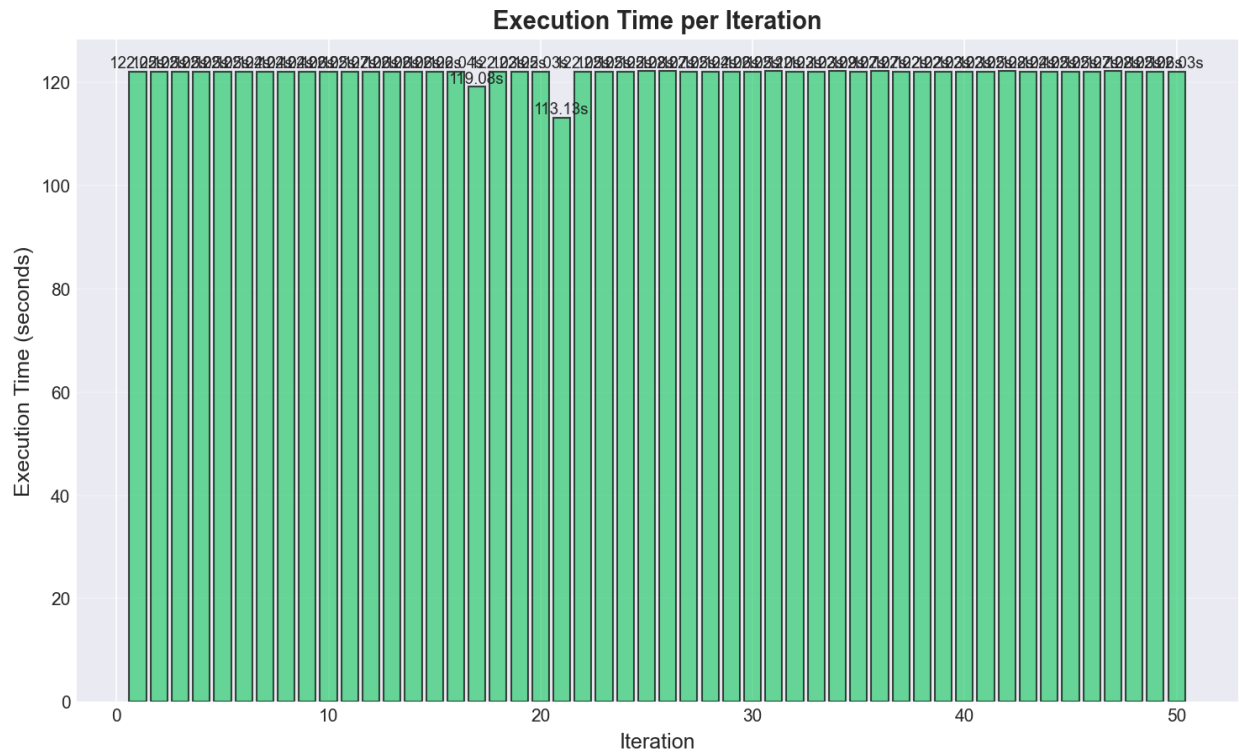
Metric	Value
Total Requests	50
Concurrent Requests	5
Model Name	llama3
Success Rate	4.0%
Throughput	0.04 req/s
Avg Execution Time	121.81s
Min Execution Time	113.13s
Max Execution Time	122.10s
Avg RAM Usage	11.35 GB
Peak RAM Usage	11.77 GB (75.1%)
Avg CPU Usage	24.7%
Peak CPU Usage	44.8%

System Information

Component	Specification
CPU	Intel64 Family 6 Model 154 Stepping 4, GenuineIntel
CPU Cores	10 cores / 12 threads
CPU Frequency	1300.0 MHz
Total RAM	15.68 GB
Platform	nt

Performance Metrics





Iteration Details

Iteration	Time (s)	Peak RAM (GB)	Avg CPU (%)	Response Length
3	122.05	11.55	28.8	0
4	122.05	11.55	29.6	0
5	122.05	11.55	29.6	0
2	122.05	11.55	28.8	0
1	122.05	11.55	29.6	0
7	122.04	11.77	30.6	0
6	122.04	11.77	29.5	0
8	122.04	11.77	29.5	0
9	122.06	11.77	33.3	0
10	122.05	11.77	30.3	0
13	122.06	11.32	15.8	0
12	122.06	11.32	15.8	0
11	122.07	11.32	15.8	0
15	122.06	11.32	16.2	0
14	122.06	11.32	16.2	0
17	119.08	11.20	23.7	2139
18	122.03	11.19	16.5	0
16	122.04	11.19	16.9	0
20	122.03	11.19	20.0	0
19	122.05	11.19	18.2	0
21	113.13	11.17	28.0	1879
23	122.05	11.15	15.4	0
22	122.05	11.15	15.4	0
24	122.05	11.15	16.7	0
25	122.08	11.15	19.2	0
26	122.07	11.25	29.4	0
28	122.04	11.20	15.0	0
27	122.05	11.20	15.0	0
29	122.06	11.20	22.5	0
30	122.05	11.20	30.6	0
31	122.10	11.32	18.9	0
33	122.03	11.27	26.4	0

32	122.03	11.27	26.4	0
34	122.09	11.27	37.8	0
35	122.07	11.27	33.3	0
36	122.07	11.43	24.4	0
37	122.02	11.43	25.0	0
38	122.02	11.43	25.0	0
39	122.03	11.43	25.0	0
40	122.03	11.43	29.3	0
41	122.05	11.33	26.4	0
43	122.04	11.32	18.6	0
42	122.08	11.32	22.9	0
44	122.05	11.32	44.8	0
45	122.05	11.32	40.3	0
46	122.07	11.31	20.0	0
48	122.05	11.32	22.4	0
47	122.08	11.32	22.4	0
50	122.03	11.31	34.2	0
49	122.06	11.31	28.7	0