

Backend Performance Pressure Test Report

1. Test Environment

Processor: 12th Gen Intel® Core™ i7-12650H

Storage: SOLIDIGM SSDPFKNU512GZ

GPU: NVIDIA GeForce RTX 4050 Laptop GPU

Memory: 16 GB DDR5

OS: Windows 11

2. Test Results Summary

Test Date: 12 Aug 2025

Test Duration: 5 minutes

Concurrent Users: 50

Endpoint	Total Requests	Successful Requests	Success Rate (%)	Avg Response Time (s)
/	2369	2369	100.0	0.177
/api/hello	3020	3020	100.0	0.191
/api/login	3573	3573	100.0	0.201
/api/start_session	1277	976	76.4	0.177
/api/get_messages/test_session_id	783	344	43.9	0.215
/api/get_sessions/test_user_id	692	356	51.4	0.257
/api/profile	636	375	59.0	0.192
/api/update_session_title	687	559	81.4	0.172
/api/delete_session	672	558	83.0	0.187
/api/search	2886	2609	90.4	0.216
/api/admin/getpdfs	981	876	89.3	0.217

/api/aichat/general	217	185	85.3	9.076
/api/aichat/rag	140	10	7.1	3.374
/api/admin/messages/need-human	683	683	100.0	0.195
/api/admin/messages/stats	1470	1470	100.0	0.176
/api/admin/messages/by-mode/general	1479	1479	100.0	0.194
/api/get_tickets	1457	1457	100.0	0.181
/api/tickets/stats	1604	1604	100.0	0.176

CPU Usage: Avg 11.4%, Max 53.4%

Memory Usage: Avg 92.4%, Max 95.5%

3. Conclusion

Most endpoints maintain **high reliability ($\geq 90\%$ success rate)**, indicating that the core system is stable.

- However, two endpoints show critical failure rates:
 - /api/aichat/rag → **7.1% success rate**
 - /api/get_messages/test_session_id → **43.9% success rate**
- This may cause by the poor hardware and parallel capabilities, which needs to be improved

Fronte Performance Pressure Test Report

1. Test Environment

Processor: 12th Gen Intel® Core™ i7-12650H

Storage: SOLIDIGM SSDPFKNU512GZ

GPU: NVIDIA GeForce RTX 4050 Laptop GPU

Memory: 16 GB DDR5

OS: Windows 11

2. Test Results Summary

Test Date: 12 Aug 2025

Concurrent Users	Success Rate (%)	Avg. Response Time (ms)	Actions per Second
1	100.00	0.10	0.89
5	100.00	0.06	4.55
10	100.00	0.07	9.50
20	100.00	0.06	17.85
50	100.00	0.06	44.97

Overall Metrics

- **Total Requests:** 2,740
- **Successful Requests:** 2,740
- **Failed Requests:** 0
- **Peak Memory Usage:** 106.55 MB
- **Total Test Duration:** 194,229.70 ms

Conclusion

Reliability maintain Success rate at 100% But have low and stable average response times by concurrency