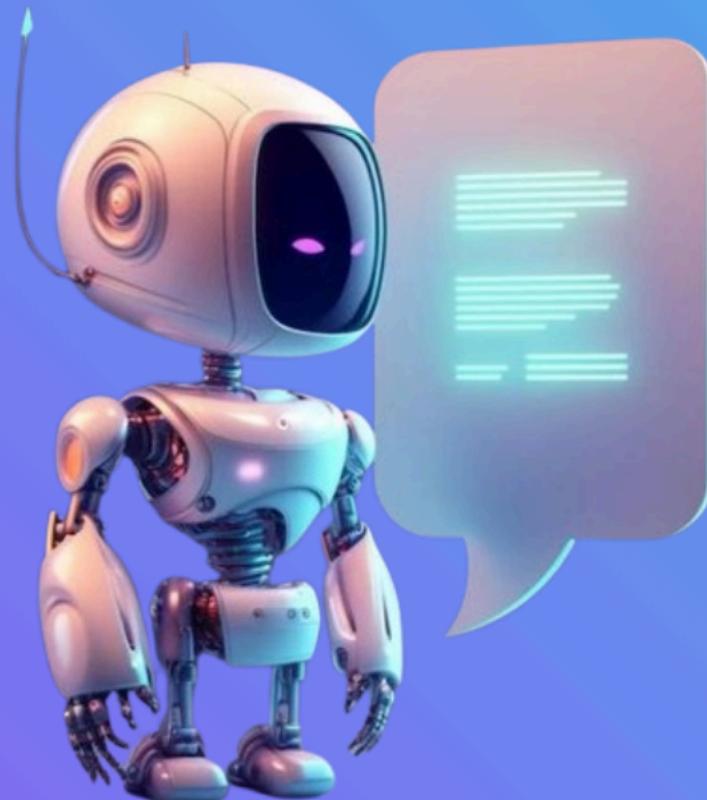


AI CHATBOT

Analysis Dashboard



 Performance Analysis

 Domain Comparison

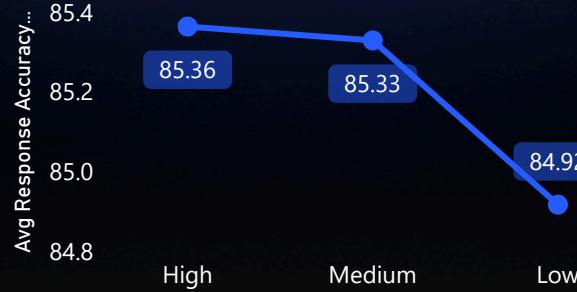
 Load Analysis

 Optimization Insights

Performance Analysis



Performance Stability Across Latency Levels



Avg Response Accuracy (%)

85.24

Avg Latency (ms)

553

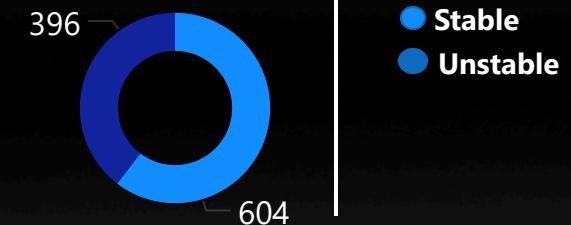
Performance Efficiency Score

21.84

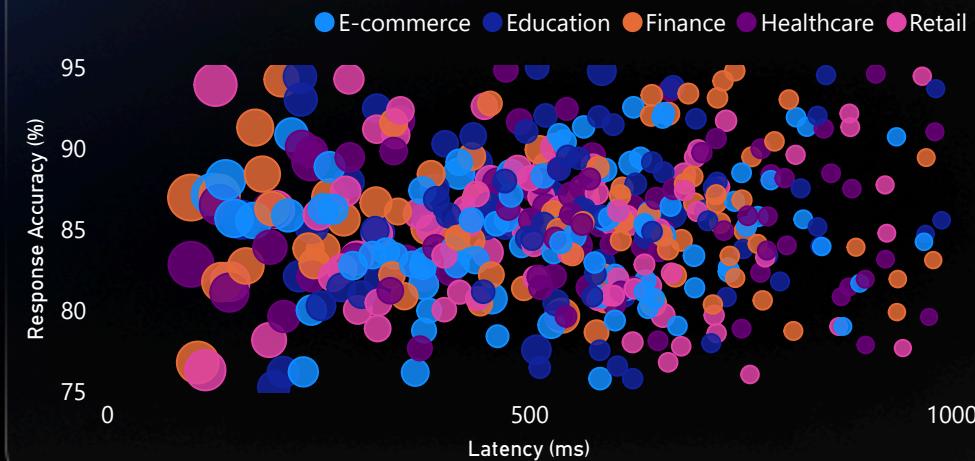
Total Unique Models

100

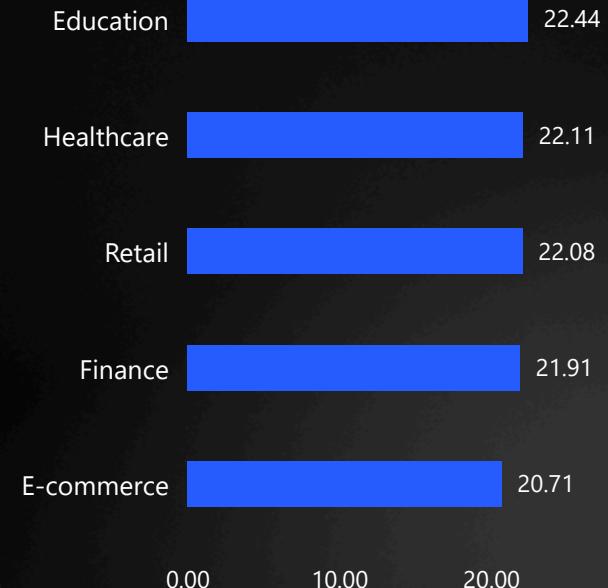
Model Performance Consistency



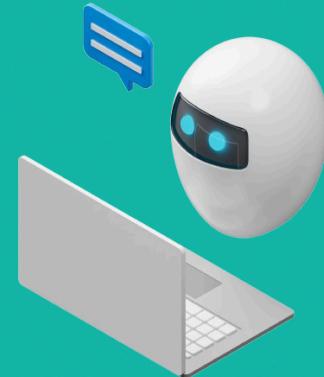
Accuracy vs Latency by Domain



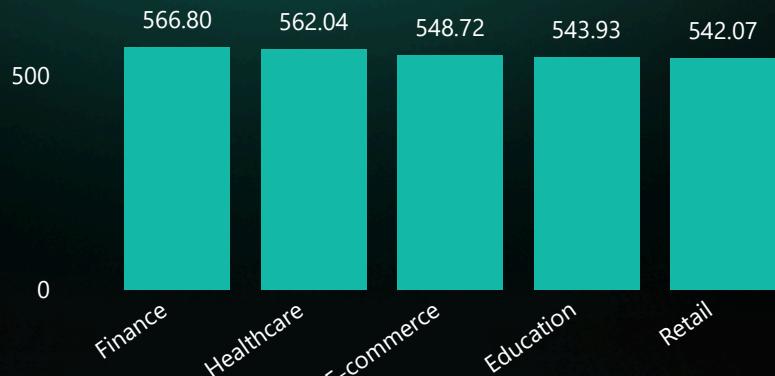
Performance Efficiency Score by Domain



Domain Comparison



Average Latency by Domain (ms)



Top Performing Domain



Education

Top Performing Domain

0.22

Top Domain Efficiency

Highest Latency Domain



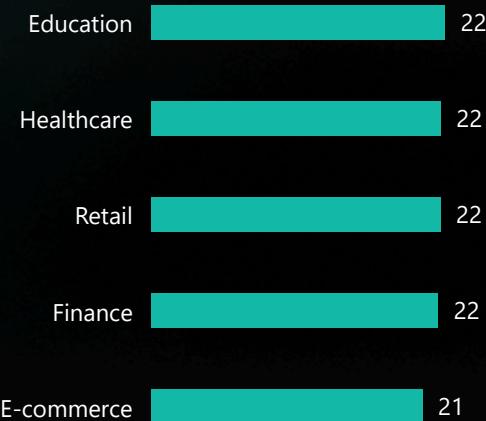
Finance

Highest Latency Domain

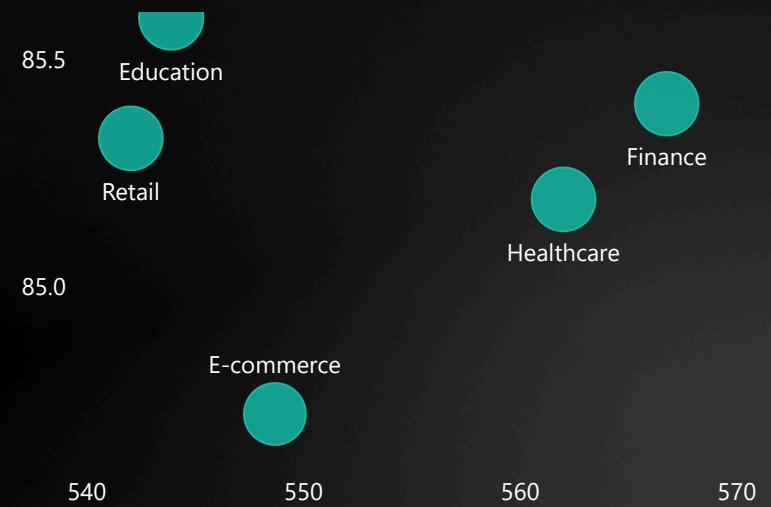
566.80

Highest Domain Latency

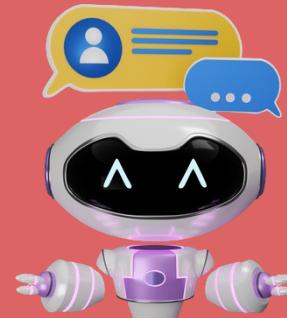
Overall Performance Efficiency by Domain



Domain-Level Accuracy vs Latency Trade-off



Load Analysis



Stress Test Pass Rate

2914.11%

Stress_Test_Pass

Accuracy Loss

-0.48

Accuracy_Degradation

High Load Success %

5306.75%

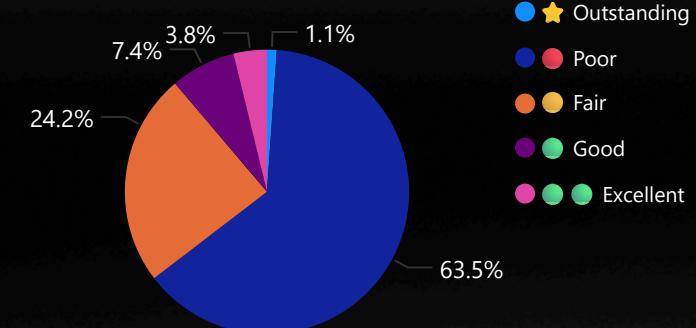
Success_Rate_HighLoad

Latency Increase %

0.02

Latency_Increase_Pct

Efficiency Rating Distribution



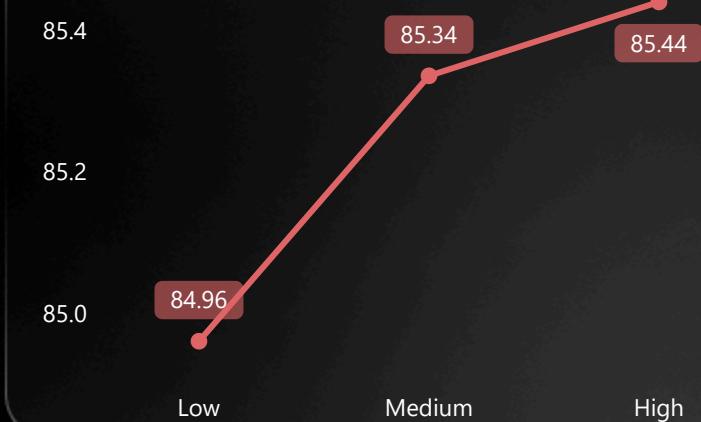
Latency Increase Under Load



Model Distribution by Efficiency Rating

efficiency_rating	Count of id
★ Outstanding	11
● Poor	635
○ Fair	242
● ● Good	74
● ● ● Excellent	38
Total	1000

Response Accuracy Across Load Levels



Optimization Insights



Most Effective Optimization Technique



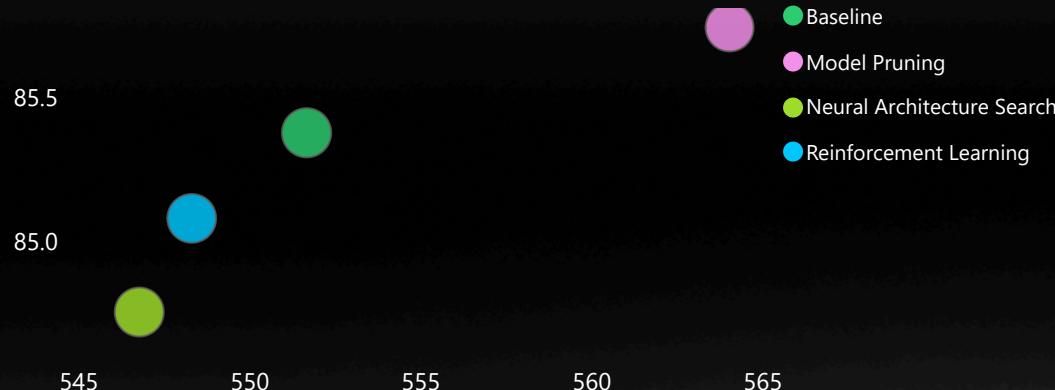
Neural Architecture Search

Best Optimization Technique

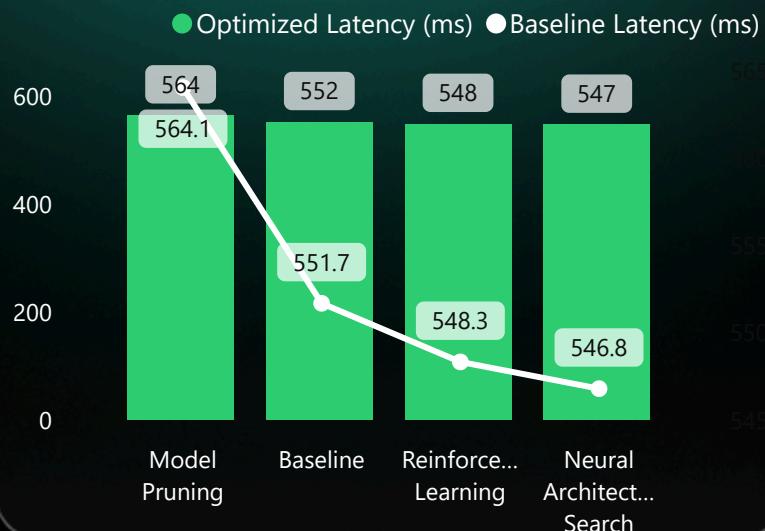
22.32%

Best Optimization Efficiency

Optimization Trade-offs: Accuracy vs Latency



Before vs After Optimization (Latency)



Best Optimization Advantage (%)

2.2

Optimization Advantage (%)

Models Needing Optimization

676

Models Needing Optimization

Performance Gain vs Baseline by Optimization Technique

