LLM Load Test Comparison (DeepInfra vs Phoenix)
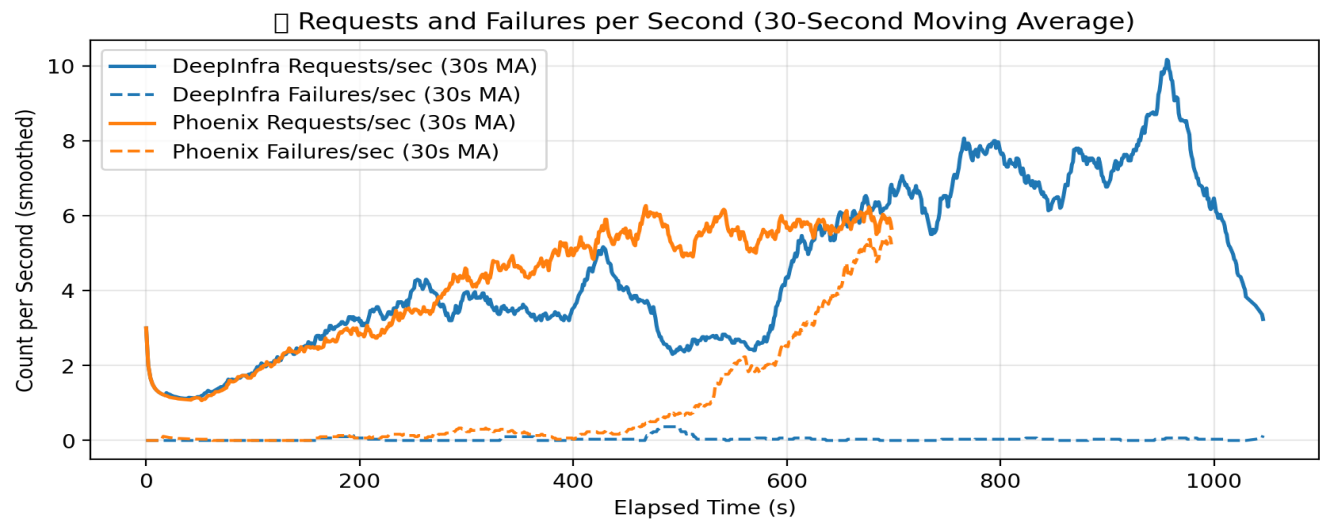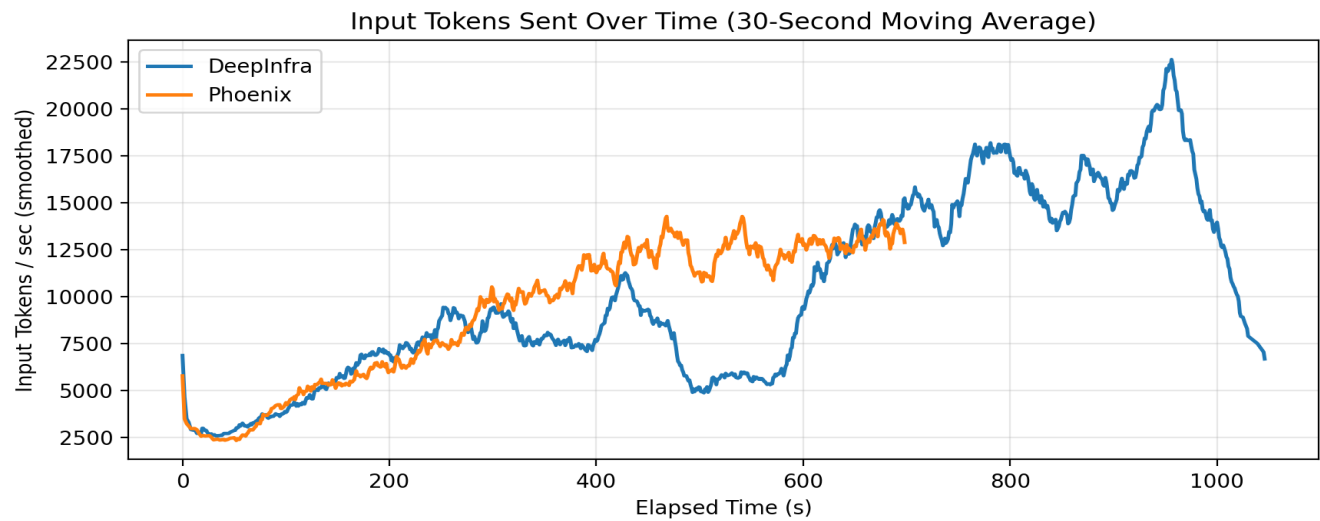
Generated on 2025-10-20 11:05:10
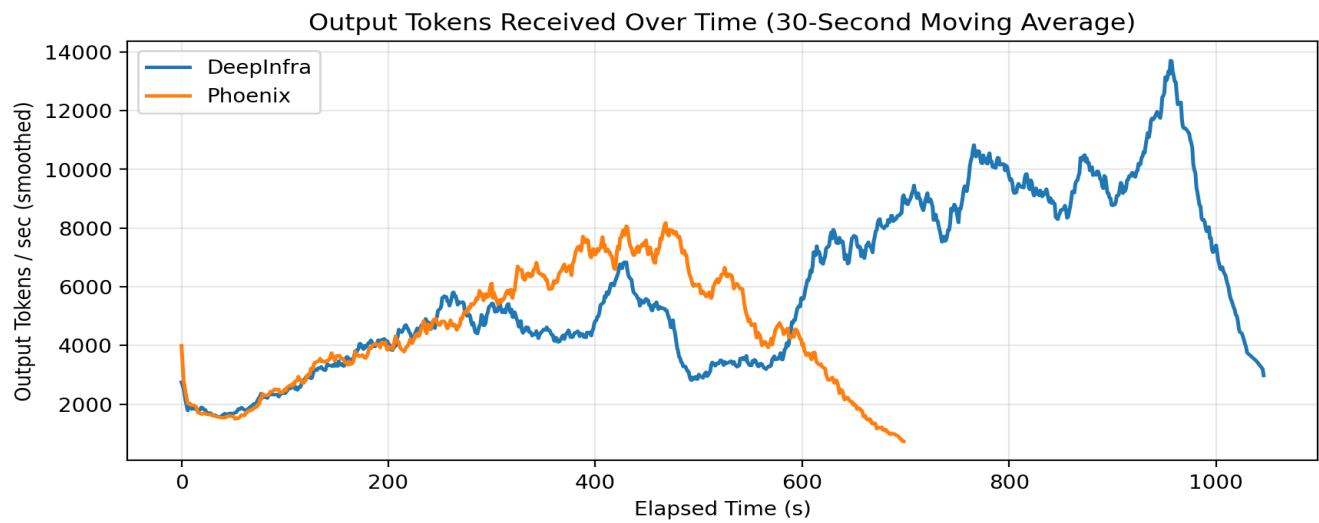
Requests and Failures per Second

Requests and Failures per Second (30-Second Moving Average)

Input Tokens Over Time

Input Tokens Sent Over Time (30-Second Moving Average)

Output Tokens Over Time

Output Tokens Received Over Time (30-Second Moving Average)

Concurrency Over Time

Concurrency Over Time (30-Second Moving Average)