

Prefill Throughput Scalability

Prefill Speed (Tokens/s)

