



Gaudi² DS LLaMA-3-8B

Gaudi2 DS LLaMA-2-7B

A100 TRT-LLM LLaMA-2-7B

H100 TRT-LLM LLaMA-3-8B

H100 TRT-LLM LLaMA-2-7B