

Prompt Length=2048, Decode Phase NPU: Op Distribution Over Model Run

