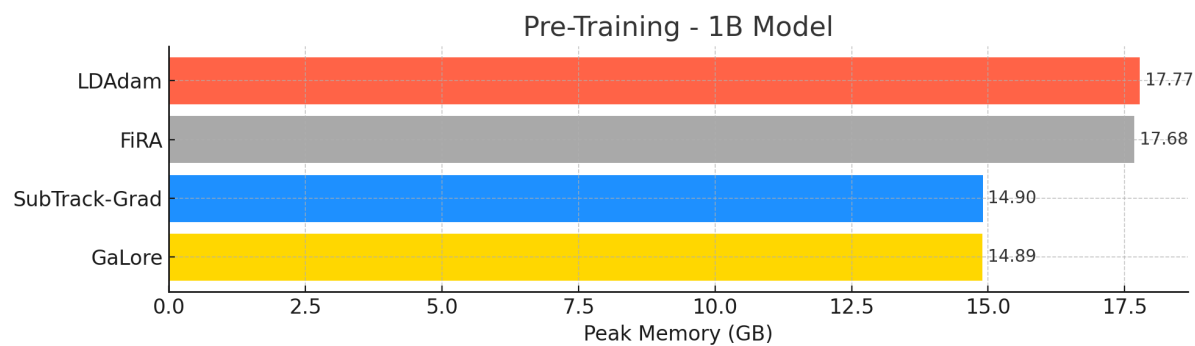


(a) Training loss of various low-rank methods during pre-training of a LLaMA-based architecture with 1B parameters, plotted against wall-time over 10K iterations.



(b) Peak Memory Consumption of various low-rank methods during pre-training of a LLaMA-based architecture with 1B parameters.