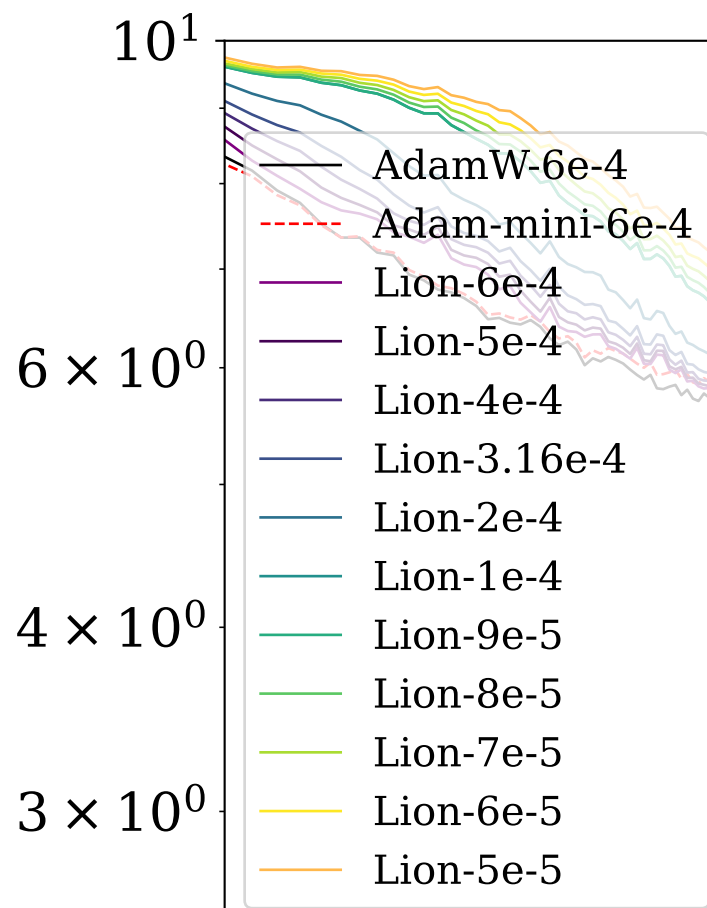


GPT2 125M Pre-train

Train Loss



Iteration