Language Model(124M) learning curve **GPT-baseline GPT** 3.4 **GPT-sincos** LLaMA\* 3.3 3.2 3.1 3.0 2.9 0 20000 40000 60000 80000 100000 iter