Language Model(350M) loss curve