Train Negative Log-Likelihood (NLL) for Single Token Generation on LM1B

Model — BD3-LM (NELBO) — BD3-LM (Tuned schedule) — AR — AR (random batch size)

