

Transformer train loss vs. batch size and number of samples, lr 0.001

