

Transformer test loss vs. batch size and number of samples

