

Transformer test loss when training last layer vs. init. magnitude

