

Transformer GOTU vs. learning rate and number of samples

