

Transformer train loss on random tasks with different lengths

