

Transformer train loss on random tasks with different numbers of templates

