

Transformer train loss with Gaussian init vs. mimetic init.

