ChimeraLM

 ${\tt dropout} = 0.1$

pooling_type = "attention"
activation = "gelu"

