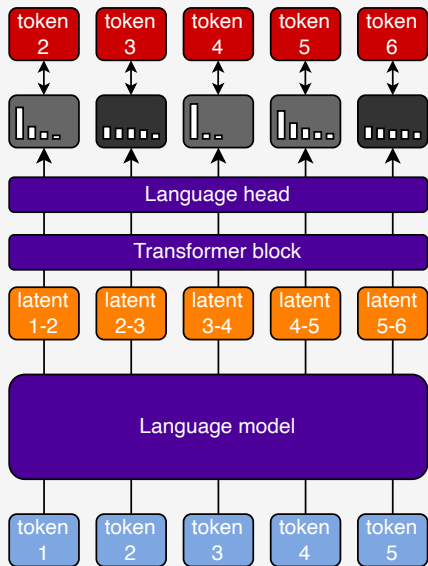


STEP 1:

Regular next-token prediction on tokens



Stop
Grad

STEP 2:

Prediction on hidden states ("latent")
at high-entropy positions

