

Transformer decoder input gradient variance

