

TRANSFORMER - Gradient Norms vs. Loss

