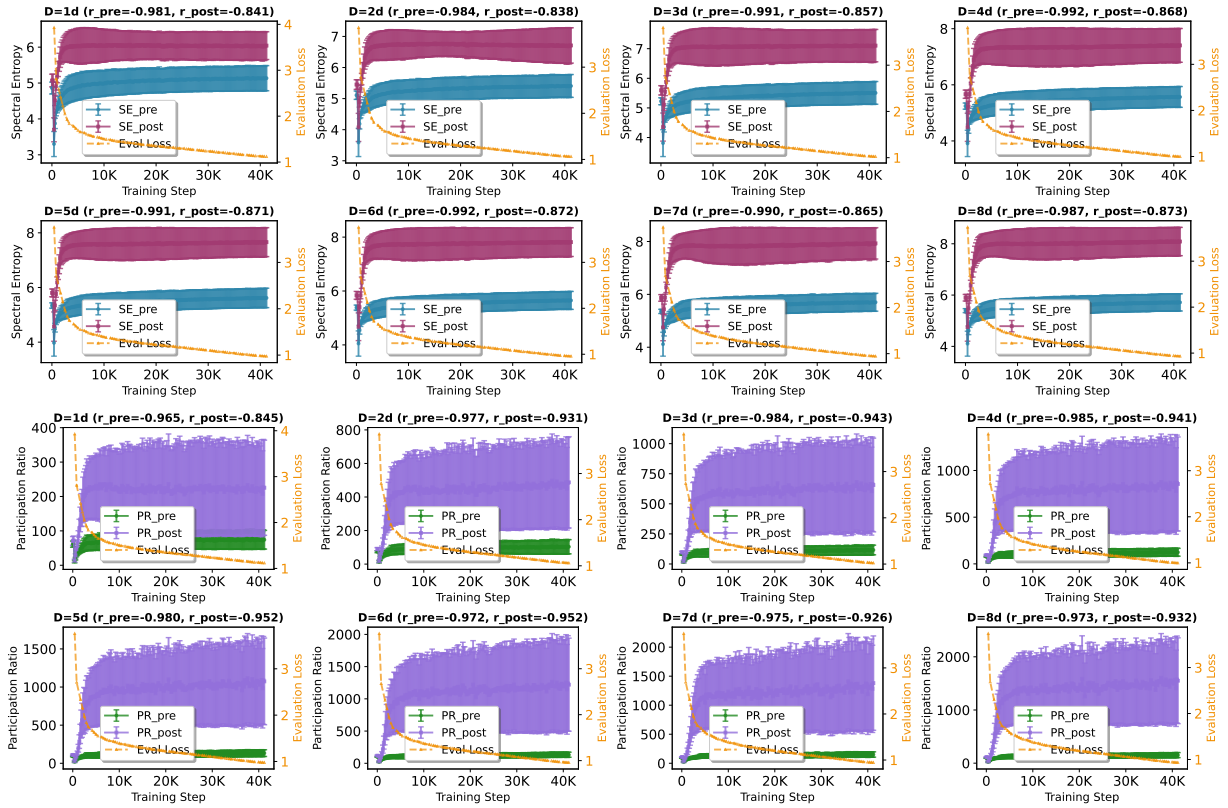

Figure 1: Layer-averaged ($\pm$SD) eigen metrics in GPT-2 (**GELU**, D=1d to 8d) with Pearson r to eval loss ($r_{\text{pre}}, r_{\text{post}}$)

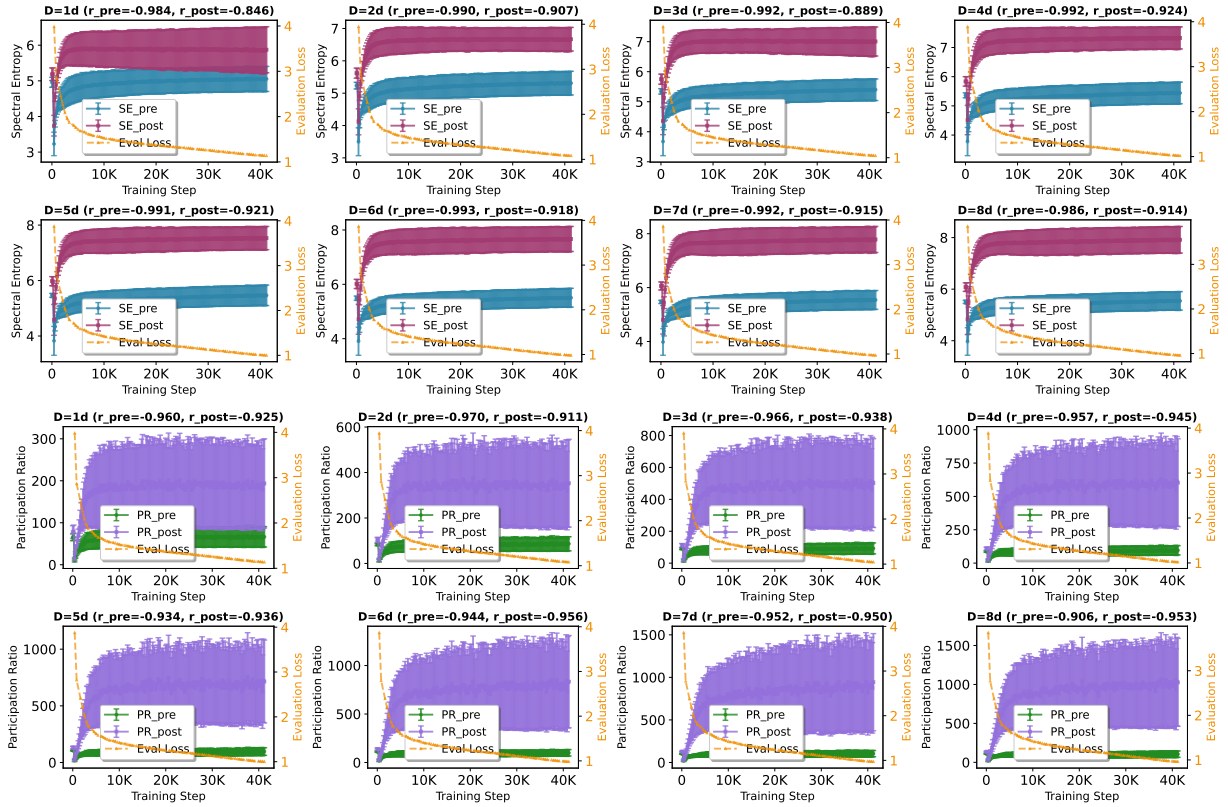

Figure 2: Layer-averaged ($\pm$SD) eigen metrics in GPT-2 (**ReLU**, D=1d to 8d) with Pearson r to eval loss ($r_{\text{pre}}, r_{\text{post}}$)

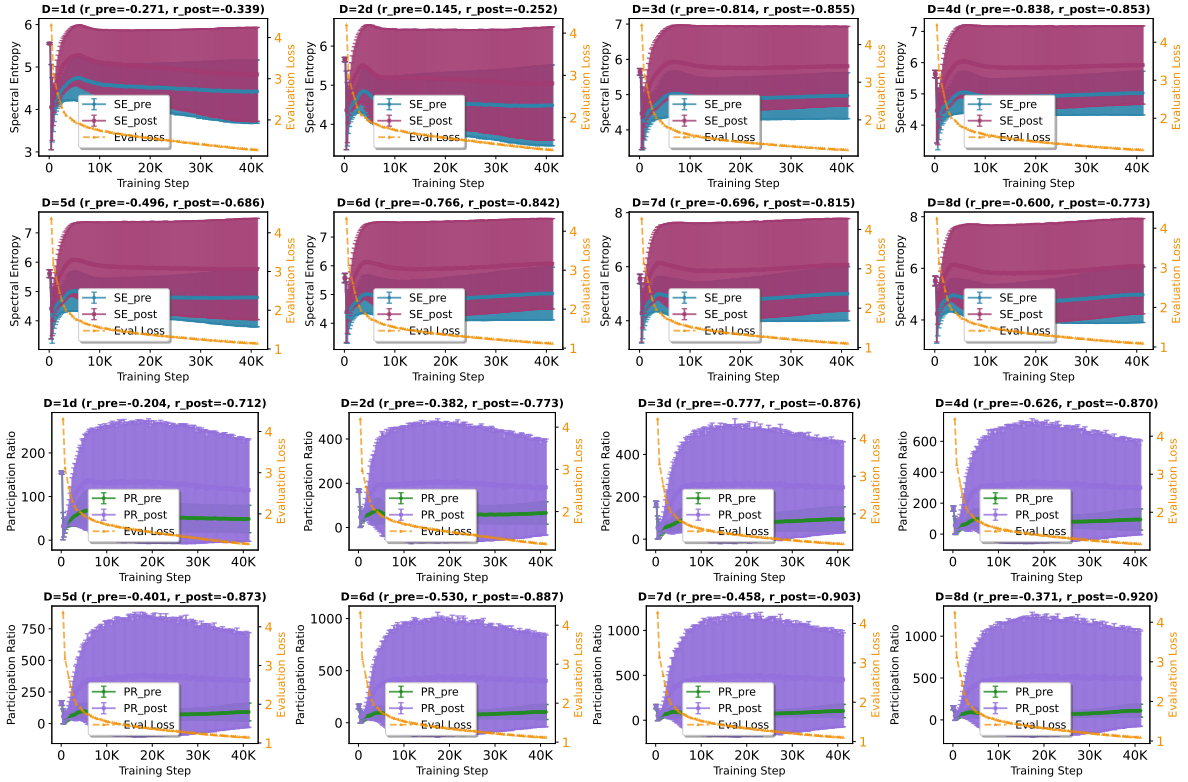

Figure 3: Layer-averaged (±SD) eigen metrics (Spectral Entropy and Participation Ratio) in **Normalization-free** GPT-2 (**GELU**, D=1d to 8d) with Pearson correlation to eval loss ($r_{\text{pre}}, r_{\text{post}}$)

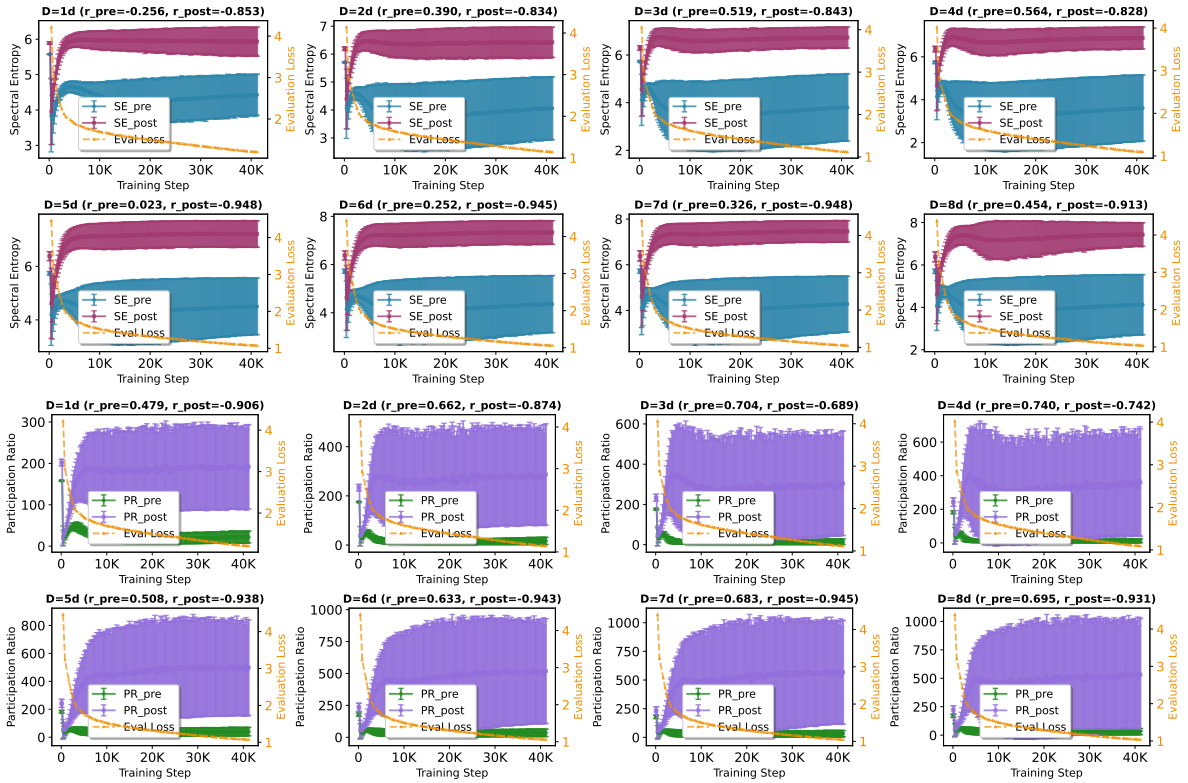

Figure 4: Layer-averaged (±SD) eigen metrics (Spectral Entropy and Participation Ratio) in **Normalization-free** GPT-2 (**ReLU**, D=1d to 8d) with Pearson correlation to eval loss ($r_{\text{pre}}, r_{\text{post}}$)

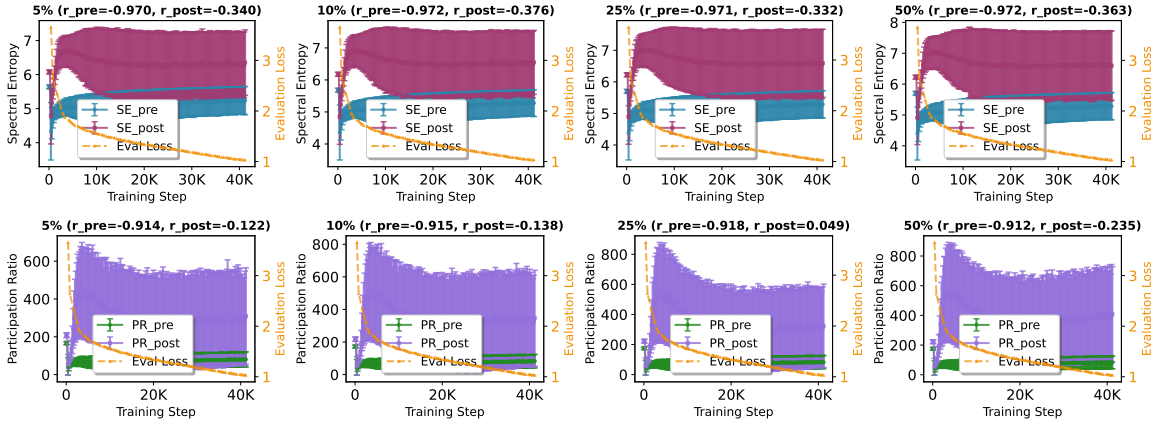

Figure 5: Layer-averaged (±SD) SE and PR in GPT-2 with Pearson r to eval loss under **sub-sampling** (5, 10, 25, 50%)

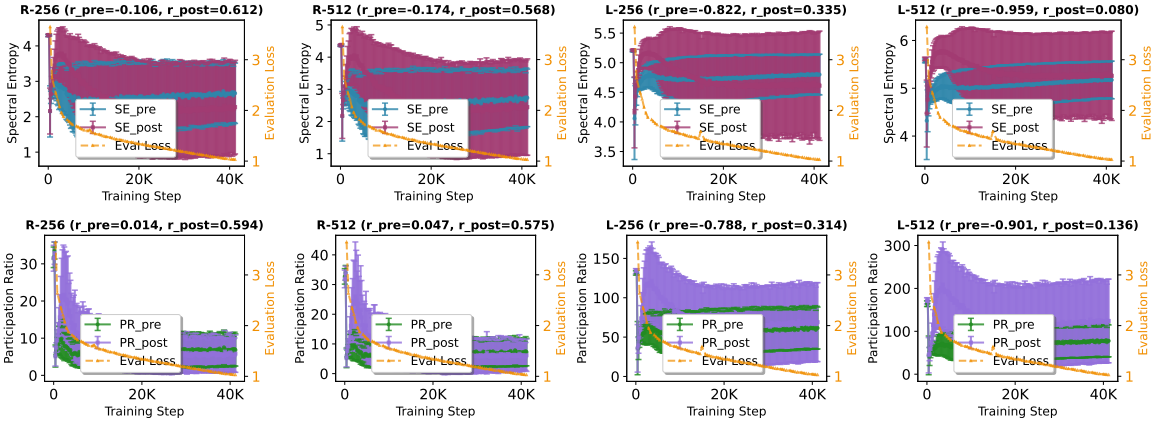

Figure 6: Layer-averaged (±SD) SE and PR in GPT-2 with Pearson r to eval loss under **low-rank approximation**.

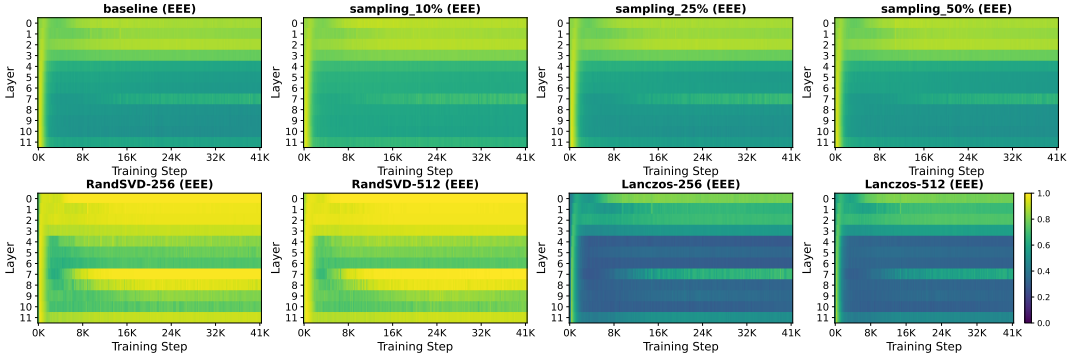

Figure 7: Impact of sampling (10%, 25%, 50%), and low-rank approximation on EEE eigenmetric in GPT-2

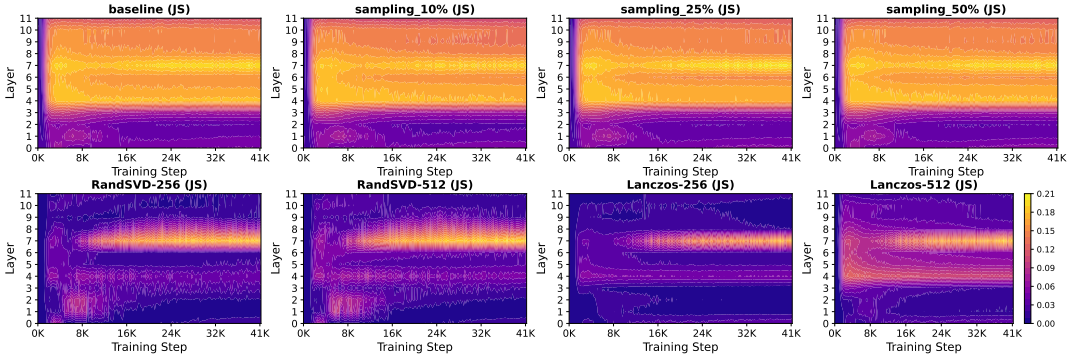

Figure 8: Impact of sampling (10%, 25%, 50%), and low-rank approximation on JS divergence eigenmetric in GPT-2

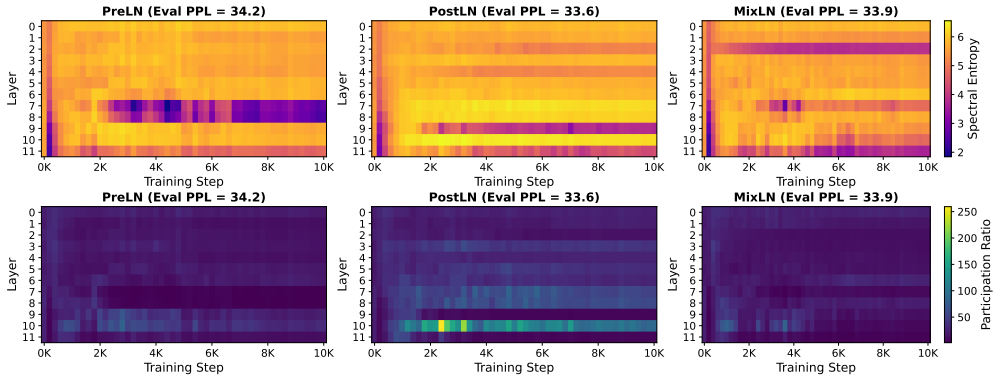

Figure 9: Eigen metrics (SE, PR) for **LLaMA-70M** PreLN, PostLN, and MixLN variants (PPL on top).

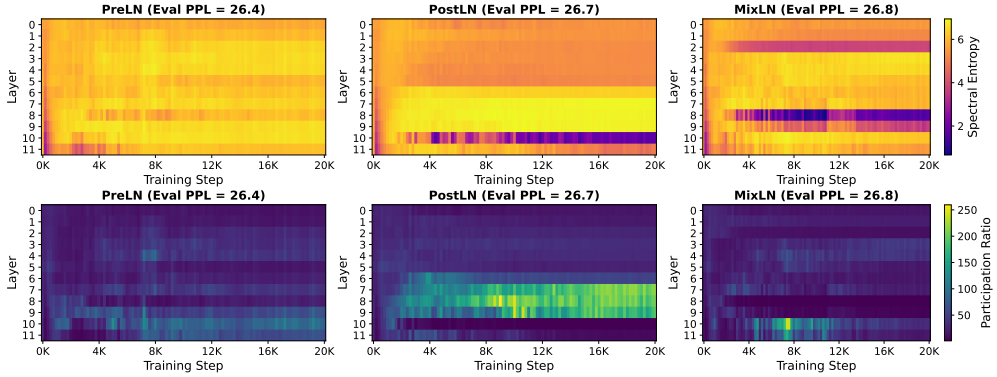

Figure 10: Eigen metrics (SE, PR) for **LLaMA-130M** PreLN, PostLN, and MixLN variants (PPL on top).

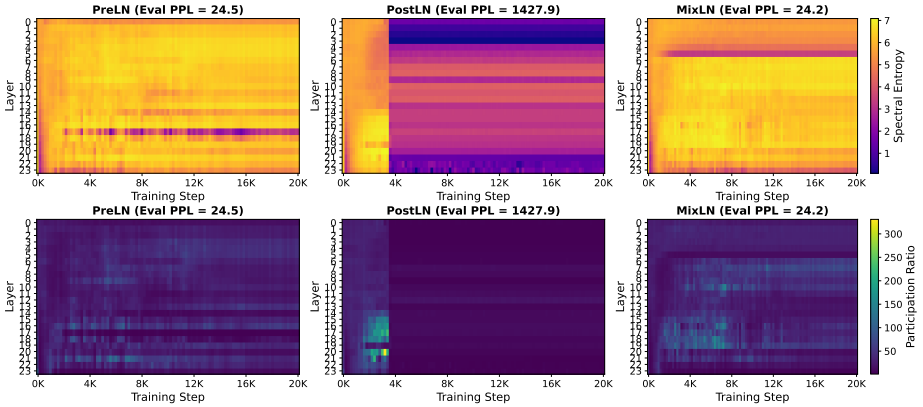

Figure 11: Eigen metrics (SE, PR) for **LLaMA-250M** PreLN, PostLN, and MixLN variants (PPL on top).

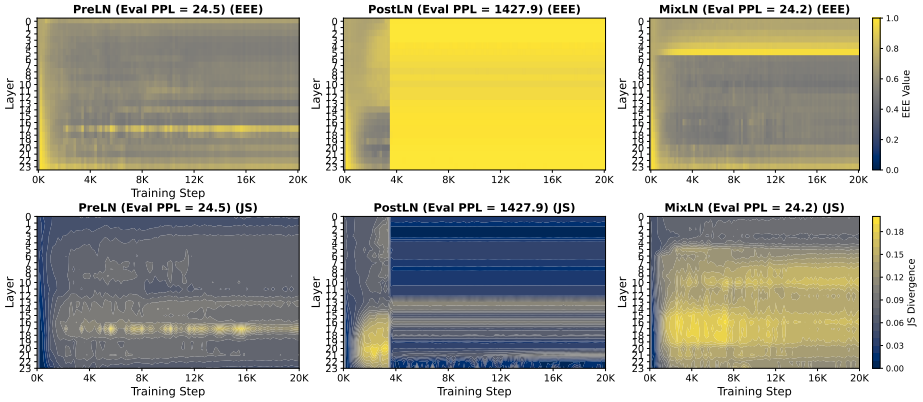

Figure 12: Eigen metrics (EEE, JS) for **LLaMA-250M** PreLN, PostLN, and MixLN variants (PPL on top)