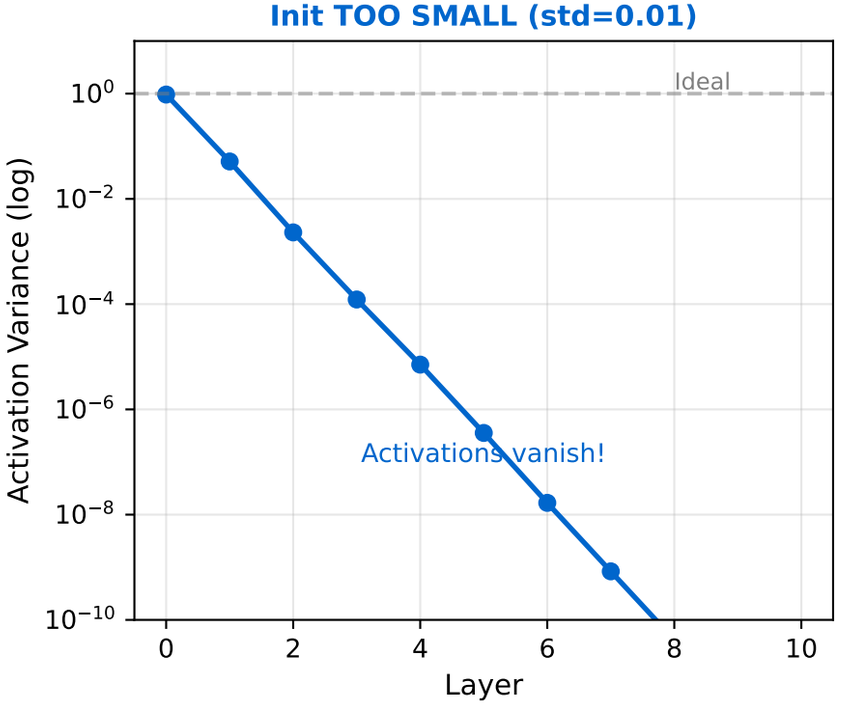
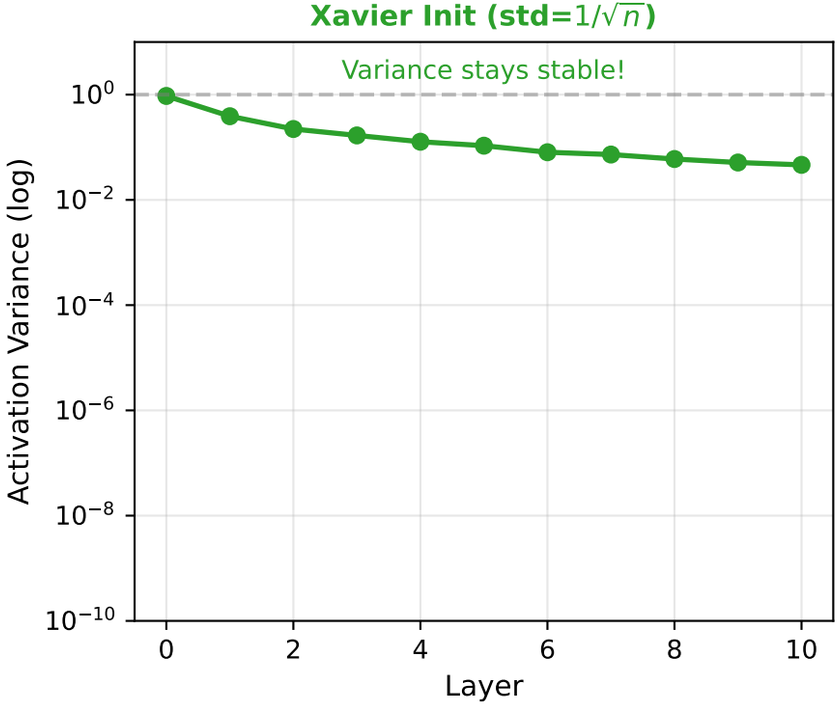
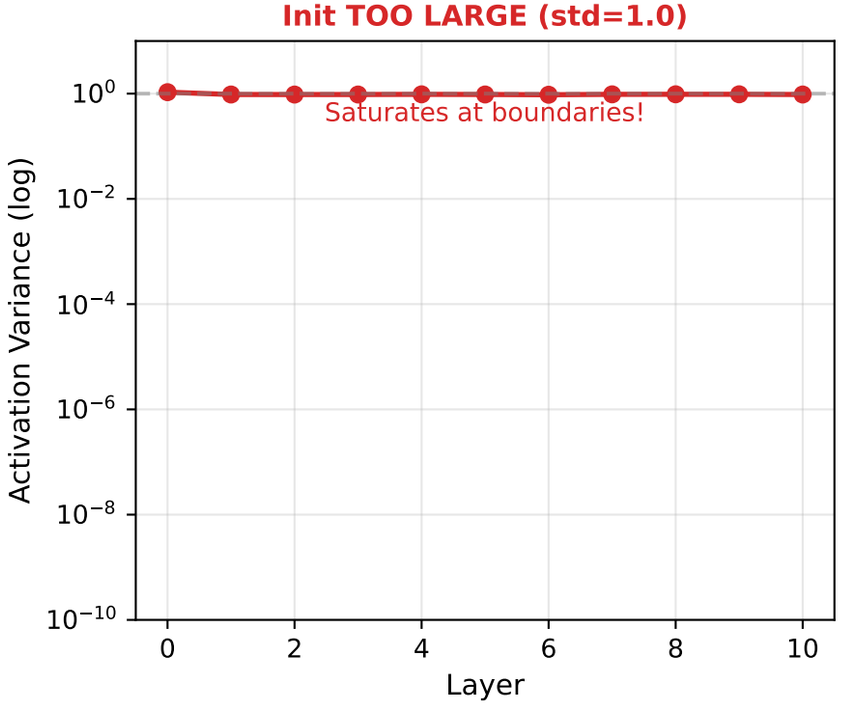
Weight Initialization: Getting It Right Matters







Xavier (Glorot): $\sigma = \sqrt{1/n_{in}}$ for tanh | He: $\sigma = \sqrt{2/n_{in}}$ for ReLU