



Dropout: A regularization technique where, during training, randomly selected neurons are ignored (their outputs are set to zero). This prevents neurons from co-adapting excessively and helps to prevent overfitting by creating a more robust model.

Residual Connection (Shortcut): A connection that skips one or more layers. In this diagram, the original input X (shown in blue) is added directly to the output of the sub-layer (Attention + Dropout). This is the "Add" step in "Add & Norm". It helps mitigate the vanishing gradient problem and allows the network to learn an identity function if needed, making it easier to train very deep networks.