

Session 4: Transformers & Attention Mechanisms

Students explore self-attention, multi-head attention, and why transformers outperform RNNs.

Activity: • Visualize attention weights • Compare transformer vs LSTM behavior