Hi there! This is a complete PyTorch implementation of the Transformer architecture from scratch, inspired by the seminal paper "Attention Is All You Need" by Vaswani et al. (2017). The implementation includes all components of the Transformer model, such as multi-head attention, positional encoding, and feed-forward networks.

Github Repository