

Transformers and Pretraining Module Suggested Readings

XCS224N Natural Language Processing with Deep Learning

1 Transformer Suggested Readings

1. [Attention Is All You Need](#)
2. [The Illustrated Transformer](#)
3. [Transformer \(Google AI blog post\)](#)
4. [Layer Normalization](#)
5. [Image Transformer](#)
6. [Music Transformer: Generating music with long-term structure](#)

2 Pre-Training Suggested Readings

1. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#)
2. [Contextual Word Representations: A Contextual Introduction](#)
3. [The Illustrated BERT, ELMo, and co.](#)
4. [Martin & Jurafsky Chapter on Transfer Learning](#)