* [Word2vec: Efficient Estimation of Word Representations in Vector Space](http://arxiv.org/pdf/1301.3781.pdf)
* [GloVe: Global Vectors for Word Representation](http://nlp.stanford.edu/pubs/glove.pdf)
* [FastText: Enriching Word Vectors with Subword Information](https://arxiv.org/pdf/1607.04606.pdf)
* [RNN: The Unreasonable Effectiveness of Recurrent Neural Networks](http://karpathy.github.io/2015/05/21/rnn-effectiveness/)
* [Sequence Modeling 1: Recurrent and Recursive Neural Nets](http://www.deeplearningbook.org/contents/rnn.html)
* [Sequence Modeling 2: Recurrent and Recursive Neural Nets](http://www.deeplearningbook.org/contents/rnn.html)
* [On the difficulty of training Recurrent Neural Networks](https://arxiv.org/pdf/1211.5063.pdf)
* [Understanding LSTM Networks](http://colah.github.io/posts/2015-08-Understanding-LSTMs/)
* [Seq2seq: Sequence to Sequence Learning with Neural Networks](https://arxiv.org/pdf/1409.3215.pdf)
* [Seq2seq with attention: Neural Machine Translation by Jointly Learning to Align and Translate](https://arxiv.org/pdf/1409.0473.pdf)
* [Transformers: Attention Is All You Need](https://arxiv.org/abs/1706.03762.pdf)
* [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/pdf/1810.04805.pdf)
* [Contextual Word Representations: A Contextual Introduction](https://arxiv.org/abs/1902.06006.pdf)
* [The Illustrated BERT, ELMo, and co.](http://jalammar.github.io/illustrated-bert/)
* [A Visual Guide to Using BERT for the First Time](http://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/)
* [GPT-2: Better Language Models and Their Implications](https://openai.com/blog/better-language-models/)