|  |  |
| --- | --- |
| Papers | Dates |
| 1. [Efficient Estimation of Word Representations in Vector Space](https://arxiv.org/pdf/1301.3781.pdf) 2. [Distributed Representations of Words and Phrases and their Compositionality](https://arxiv.org/pdf/1310.4546.pdf) 3. [GloVe: Global Vectors for Word Representation](https://nlp.stanford.edu/pubs/glove.pdf) 4. [Enriching Word Vectors with Subword Information](https://arxiv.org/pdf/1607.04606.pdf) | August 8, 10 |
| 1. [Extensions of Recurrent Neural Network Language Model](http://www.fit.vutbr.cz/research/groups/speech/publi/2011/mikolov_icassp2011_5528.pdf) 2. [Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation](https://arxiv.org/pdf/1406.1078v3.pdf) 3. [Sequence to Sequence Learning with Neural Networks](https://arxiv.org/pdf/1409.3215.pdf) 4. [Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling](https://arxiv.org/pdf/1412.3555.pdf) 5. [Convolutional Neural Networks for Sentence Classification](https://aclanthology.org/D14-1181.pdf) 6. [Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks](https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/43455.pdf) | August 22, 24, 29  September 5 |
| 11. [An overview of Gradient Descent Optimization Algorithms](https://arxiv.org/pdf/1609.04747.pdf)  12. [Overview of mini-batch Gradient Descent](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf)  13. [Understanding the difficulty of Training Deep Feedforward Neural Networks](https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf)  14. [Cyclical Learning Rates for Training Neural Networks](https://arxiv.org/pdf/1506.01186.pdf)  15. [Don’t Decay the Learning Rate, Increase the Batch Size](https://openreview.net/forum?id=B1Yy1BxCZ) | September 12, 14 |
| 16. [Google’s Machine Translation..](https://arxiv.org/pdf/1609.08144v2.pdf).  17. [Neural Machine Translation of Rare Words with Subword Units](https://aclanthology.org/P16-1162.pdf) | October 3 |
| 18. [Attention is All You Need](https://arxiv.org/abs/1706.03762)  19. [Nystromformer: A Nystrom-Based Algorithm for Approximating Self-Attention](https://arxiv.org/abs/2102.03902)  20. [Linformer: Self-Attention with Linear Complexity](https://arxiv.org/pdf/2006.04768v3.pdf)  21. [You may not need Attention](https://arxiv.org/pdf/1810.13409.pdf)  21. [Attention is not Explanation](https://arxiv.org/pdf/1902.10186.pdf)  22. [Attention is not not Explanation7](https://aclanthology.org/D19-1002.pdf) | October 10, 12, 17 |
| 23. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/pdf/1810.04805)  24. [RoBERTa: A Robustly Optimized BERT Pretraining Approach](https://arxiv.org/pdf/1907.11692)  25. [DeBERTAa: Decoding-enhanced BERT with Disentangled Attention](https://arxiv.org/pdf/2006.03654)  26. [BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation and Comprehension](https://arxiv.org/pdf/1910.13461v1)  27. [DistilBERT, a distilled version of BERT](https://arxiv.org/pdf/1910.01108)  28. [Large Batch Optimization for Deep Learning](https://arxiv.org/pdf/1904.00962) | October 26, 31, Nov 2 |
| 29. [Improving Language Understanding by Generative Pretraining](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)  30. [Language Models are Unsupervised Multitask Learners](https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)  31. [Language Models are Few-Shot Learners](https://arxiv.org/pdf/2005.14165) | Nov 7, 9 |
| 32. [Graph Convolutional Networks for Text Classification](https://arxiv.org/pdf/1809.05679)  33. [Graph Attention Networks](https://arxiv.org/pdf/1710.10903.pdf)  34. [Graph Transformer Networks](https://arxiv.org/pdf/1911.06455.pdf)  35. [GraphBERT: Only Attention is needed for Learning Graph Representations](https://arxiv.org/pdf/2001.05140.pdf)  36. [GPT-GNN: Generative Pretraining of Graph Neural Networks](https://arxiv.org/pdf/2006.15437.pdf)  37. [Plug and Play Language Models](https://arxiv.org/pdf/1912.02164.pdf) | Nov 14, 16, 21, 23 |
|  |  |