

Textbooks:

1. [Deep Learning](#), by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
 - 5 - Machine Learning Basics
 - 6 - Deep Feedforward Networks
 - 7 - Regularization for Deep Learning
 - 8 - Optimization for Training Deep Models
 - 10 - Sequence Modeling: Recurrent and Recursive Nets
2. [Speech and Language Processing \(3rd edition\)](#), by Daniel Jurafsky and James H. Martin
 - 3 - N-gram Language Models
 - 6 - Vector Semantics and Embeddings
 - 7 - Neural Networks and Neural Language Models
 - 9 - RNNs and LSTMs
 - 10 - Transformers and Pretrained Language Models
 - 11 - Fine-tuning and Masked Language Modeling

Background papers:

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. 2017. [Attention Is All You Need](#), In *Advances in Neural Information Processing Systems (NeurIPS)*.
2. Alec Radford and Karthik Narasimhan. 2018. [Improving Language Understanding by Generative Pre-Training](#).
3. Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.
4. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li and Peter J. Liu. 2020. [Exploring the limits of transfer learning with a unified text-to-text transformer](#). In *The Journal of Machine Learning Research*, 21.
5. Kevin Clark, Minh-Thang Luong, Quoc V. Le and Christopher D. Manning. 2020. [ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators](#). In *International Conference on Learning Representations*
6. Yoav Levine, Barak Lenz, Opher Lieber, Omri Abend, Kevin Leyton-Brown, Moshe Tennenholtz and Yoav Shoham. 2021. [PMI-Masking: Principled masking of correlated spans](#). In *International Conference on Learning Representations*.
7. Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang and Samuel R. Bowman. 2020. [BLiMP: The Benchmark of Linguistic Minimal Pairs for English](#). In *Transactions of the Association for Computational Linguistics, Volume 8*.

8. Timo Schick and Hinrich Schütze. 2021. [Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*.
9. Tianyu Gao, Adam Fisch and Danqi Chen. 2021. [Making Pre-trained Language Models Better Few-shot Learners](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.
10. Nikunj Saunshi, Sadhika Malladi and Sanjeev Arora. 2021. [A Mathematical Exploration of Why Language Models Help Solve Downstream Tasks](#). In *International Conference on Learning Representations*.