

1. Textbooks
 - a. [Speech and Language Processing](#) by Jurafsky and Martin, Jan 2025 edition - chapters 2-12:
 - i. Regular Expressions, Tokenization, Edit Distance
 - ii. N-gram Language Models
 - iii. Naive Bayes, Text Classification, and Sentiment
 - iv. Logistic Regression
 - v. Vector Semantics and Embeddings
 - vi. Neural Networks
 - vii. RNNs and LSTMs
 - viii. Transformers
 - ix. Large Language Models
 - x. Masked Language Models
 - xi. Model Alignment, Prompting, and In-Context Learning
2. Classic papers
 - a. [Attention is all you need](#)
 - b. [Sequence to Sequence Learning with Neural Networks](#)
 - c. [REINFORCE](#)
3. Pruning
 - a. [Autencoding Variational Bayes](#)
 - b. [The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables](#)
 - c. [Learning Sparse Neural Networks through L0 Regularization](#)
 - d. [Structured Pruning of Large Language Models](#)
4. Interpretability
 - a. [A Mathematical Framework for Transformer Circuits](#)
 - b. [In-context Learning and Induction Heads](#)
 - c. [Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets](#)
 - d. [Learning Transformer Programs](#)
 - e. [Towards Automated Circuit Discovery for Mechanistic Interpretability](#)