

Additional resources for the “Introduction to Language Models” lecture of the course IE686 Large Language Models and Agents at the University of Mannheim HWS2025:

1. Language Models and Processing (Textbook):

Daniel Jurafsky & James H. Martin: [Speech and Language Processing](#). (3rd edition draft)

2. Word2Vec and fastText:

Illustration/Tutorial: <https://jalammar.github.io/illustrated-word2vec/>

fastText word vectors and docs: <https://fasttext.cc/docs/en/support.html>

gensim python library for word2Vec and fastText:

<https://radimrehurek.com/gensim/models/word2vec.html>

Papers:

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. [Distributed Representations of words and Phrases and their Compositionality](#). *Advances in Neural Information Processing Systems*, 26.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T., 2017. [Enriching word vectors with subword information](#). *Transactions of the Association for Computational Linguistics*, 5, pp.135-146.

3. The Transformer architecture:

Videos:

<https://www.youtube.com/watch?v=wjZofJX0v4M>

<https://www.youtube.com/watch?v=eMlx5fFNoYc>

Illustration/Tutorial:

<https://jalammar.github.io/illustrated-transformer/>

<https://huggingface.co/learn/nlp-course/chapter1/4?fw=pt>

Attention Visualization Demo: <https://github.com/jessevig/bertviz>

Paper: Vaswani, A., et al., 2017. [Attention is All You Need](#). *Advances in Neural Information Processing Systems*.

4. The GPT Transformers:

Illustration/Tutorial: <https://jalammar.github.io/illustrated-gpt2/>

Model Demo: <https://huggingface.co/openai-community/gpt2>

Papers:

Radford et al., 2018. [Improving Language Understanding by Generative Pre-Training](#)

Radford et al., 2019. [Language Models are Unsupervised Multitask Learners](#)

Brown et al., 2020. [Language Models are Few-shot Learners](#). In Proceedings of the 34th International Conference on Neural Information Processing Systems (pp. 1877-1901).

5. The BERT Transformer and its variants:

Illustration/Tutorial: <https://jalanmar.github.io/illustrated-bert/>

Model Demo: <https://huggingface.co/google-bert/bert-base-cased>

Papers:

Devlin et al., 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186).

Liu et al., 2019. [RoBERTa: A Robustly Optimized BERT Pretraining Approach](#). *arXiv preprint arXiv:1907.11692*.

Sanh, V., 2019. [DistilBERT, A Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter](#). *arXiv preprint arXiv:1910.01108*.

Clark et al., 2020. [ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators](#). In *International Conference on Learning Representations*.

6. The T5 Transformer:

Model Demo: <https://huggingface.co/google-t5/t5-base>

Paper: Raffel et al., 2020. [Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer](#). *Journal of Machine Learning Research*, 21(140), pp.1-67.

7. The BART Transformer

Model Demo: <https://huggingface.co/facebook/bart-base>

Paper: Lewis et al., 2020. [BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7871-7880).

8. Survey: History of Pre-trained Language Models:

Zhou et al.: [A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT](#). 2023. *arXiv:2302.09419*.