* [BERT Research Series](https://www.youtube.com/playlist?list=PLam9sigHPGwOBuH4_4fr-XvDbe5uneaf6) (Concept, code explanation, notebook code)
* <https://huggingface.co/transformers/>
* <https://nlp.seas.harvard.edu/2018/04/03/attention.html>
* <http://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/>
* <https://medium.com/@_init_/why-bert-has-3-embedding-layers-and-their-implementation-details-9c261108e28a>
* <https://kazemnejad.com/blog/transformer_architecture_positional_encoding/>
* <https://timodenk.com/blog/linear-relationships-in-the-transformers-positional-encoding/>
* <https://medium.com/huggingface/distilbert-8cf3380435b5>
* <https://blog.rasa.com/compressing-bert-for-faster-prediction-2/>
* [Kaggle Coffee Chat: Jacob Devlin (Google Researcher, BERT author) | Kaggle](https://www.youtube.com/watch?v=u91645MFytY)