# The Attention Mechanism – Review

D. Gueorguiev 1/7/2023

## Introductory Notes

The attention mechanism is a part of a neural architecture that enables dynamically to select relevant features in the input data which in NLP is typically a sequence of textual elements. The idea behind attention is to compute a weight distribution on the input sequence, assigning higher values to more relevant elements.

## Appendix

### Bi-directional Neural Networks

*Problem Statement*

Consider a (time) sequence of input data vectors

and a sequence of corresponding output data vectors

with neighboring data-pairs in time being statistically independent. Given time sequences and as training data, the aim is to learn the rules to predict output data given the input data. Inputs and outputs, can, in general, be continuous and/or categorical variables.

## Bibliography

[1] [Attention in Natural Language Processing, Andrea Galassi, Marco Lippi, Paolo Torroni, 2019](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/AttentionInNaturalLanguageProcessing.pdf)

[2] [Bidirectional Recurrent Neural Networks, Mike Schuster, Kuldip Paliwal, IEEE Transactions on Signal Processing, 1997](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Bidirectional_Recurrent_Neural_Networks_Schuster_Paliwal_1997.pdf)

[3] [Neural Machine Translation by Jointly Learning To Align and Translate, Dzmitry Bahdanau, K. Cho, Yoshua Bengio, 2016](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/NeuralMachineTranslationByJointlyLearningToAlignAndTranslateBahdanau2015.pdf)

[4] [Recurrent Continuous Translation Models, Nal Kalchbrenner, Phil Blunsom, Oxford U., 2013](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Recurrent_continuous_translation_models_Kalchbrenner_Blunsson_OxfordU_2013.pdf)

[5] [On The Properties of Neural Machine Translation: Encoder-Decoder Approaches, K. Cho, B. van Merrienboer, Dzmitry Bahdanau, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/OnthePropertiesOfNeuralMachineTranslationEncoderDecoderApproaches.pdf)

[6] [Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, K. Cho, B. van Merrienboer, C. Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/LearningPhraseRepresentationsUsingRNNEncoderDecoderForStatisticalMachineTranslation.pdf)

[7] [Autoencoders, Dor Bank, Noam Koenigstein, Raja Giryes, 2021](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Autoencoders.pdf)

[8] [Attention and Augmented Recurrent Neural Networks, Chris Olah, Shan Carter, Google Brain blog, 2016](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/AttentionandAugmentedRecurrentNeuralNetworks.pdf)

[9] [Attention is All You Need, Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit et al, NIPS, 2017](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Attention-is-all-you-need-NIPS-2017.pdf)

[10] [Transforming Auto-encoders, G.E. Hinton, A. Krizhevsky, S.D. Wang, U of Toronto, 2010](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/TransformingAutoencodersHinton.pdf)

[11] [Sequence to Sequence Learning with Neural Networks, Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Google Research, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/SequencetoSequenceLearningwithNeuralNetworksSutsekver2014.pdf)

[12] [Generating Sequences With Recurrent Neural Networks, Alex Graves, U of Toronto, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Generating_Sequences_With_Recurrent_Neural_Networks_Graves_2014.pdf)

[13] [A Tutorial on Training Recurrent Neural Networks, covering BPPT, RTRL, EKF and the “echo state network” approach, Herbert Jaeger, Fraunhofer Institute for Autonomous Intelligent Systems, 2003](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/TutorialOnRNNAndBPTTJaeger2002.pdf)

[14] [Understanding LSTM: a Tutorial into Long Short-Term Memory Recurrent Neural Networks, Ralf C. Staudemeyer, Eric Rothstein Morris, 2019](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/TutorialOnLongShortTermMemory2019.pdf)

[15] [Long Short-Term Memory, Sepp Hochreiter, Juergen Schmidhuber, Neural Computation 9(8), 1997](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/LongShortTermMemory.pdf)

[16] [Sequence Modeling With Neural Networks (Part 1): Language & Seq2Seq, Nathan Lintz’s blog, 2016](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/SequenceModelingWithNeuralNetworksPart1_Seq2Seq.pdf)

[17] [Sequence Modeling With Neural Networks (Part 2): Attention Models, Nathan Lintz’s blog, 2016](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/SequenceModelingwithNeuralNetworksPart2_AttentionModels.pdf)