



### 9.3 Дополнительные материалы

Вы прошли больше 80% курса, оставьте отзыв

[Оставить отзыв](#)    [Нет, спасибо](#)

Здесь собраны ссылки на дополнительные материалы к модулям "Attention" и "Transformers".

- Увлекательная статья об истории решения задачи машинного перевода с 1950х годов до наших дней: <https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5/>
- Статья "Neural Machine Translation by Jointly Learning to Align and Translate" (статья, в которой был впервые представлен механизм Attention): <https://arxiv.org/pdf/1409.0473.pdf>
- Статья "Attention is all you need" (статья, в которой был представлен Transformer для решения задачи машинного перевода): <https://arxiv.org/pdf/1706.03762.pdf>
- Визуальное объяснение position encodings: зачем они нужны и как их построить: <https://www.youtube.com/watch?v=dichIcUZfOw>
- Очень подробная статья про position encodings (видео по ссылке выше основано на этой статье). Статья объясняет, почему position encodings в той форме, что мы рассматривали на лекции, помогает механизму Attention получать информацию об относительной позиции токенов в предложении: [https://kazemnejad.com/blog/transformer\\_architecture\\_positional\\_encoding/](https://kazemnejad.com/blog/transformer_architecture_positional_encoding/)
- Статья с визуальным объяснением архитектуры Transformer (хороший вариант, чтобы посмотреть на объяснение архитектуры с другого боку, кроме той, что была на лекции): <http://jalammar.github.io/illustrated-transformer/>
- Создание своего Трансформера с примерами кода: <https://towardsdatascience.com/build-your-own-transformer-from-scratch-using-pytorch-84c850470dcb>
- Статья с объяснением следующего явления: если добавить слой LayerNorm перед слоем Attention, а не после, то LayerNorm помогает механизму Attention сформировать такой вектор query из эмбединга, что все векторы key лежат на одинаковом расстоянии от вектора query, т.е., другими словами, становятся одинаково доступны. Это делает механизм внимания более эффективным: <https://lessw.medium.com/what-layernorm-really-does-for-attention-in-transformers-4901ea6d890e#:~:text=Normalization%20via%20LayerNorm%20has%20been.and%20gradients%20on%20the%20backward.>
- Есть много инструментов визуализации весов attention. Один из них — BertViz. Эта статья объясняет, как им пользоваться: <https://www.comet.com/site/blog/explainable-ai-for-transformers/>

 12 



#### Комментарии

Будьте вежливы и соблюдайте наши [принципы сообщества](#). Пожалуйста, не оставляйте решения и подсказки в комментариях, для этого есть отдельный форум.

[Оставить комментарий](#)

[12.1 Урок для пересдачи Д..](#)

