Quick review of attention:

Fig. 4. The encoder-decoder model with additive attention mechanism in *Bahdanau et al., 2015*.

https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html
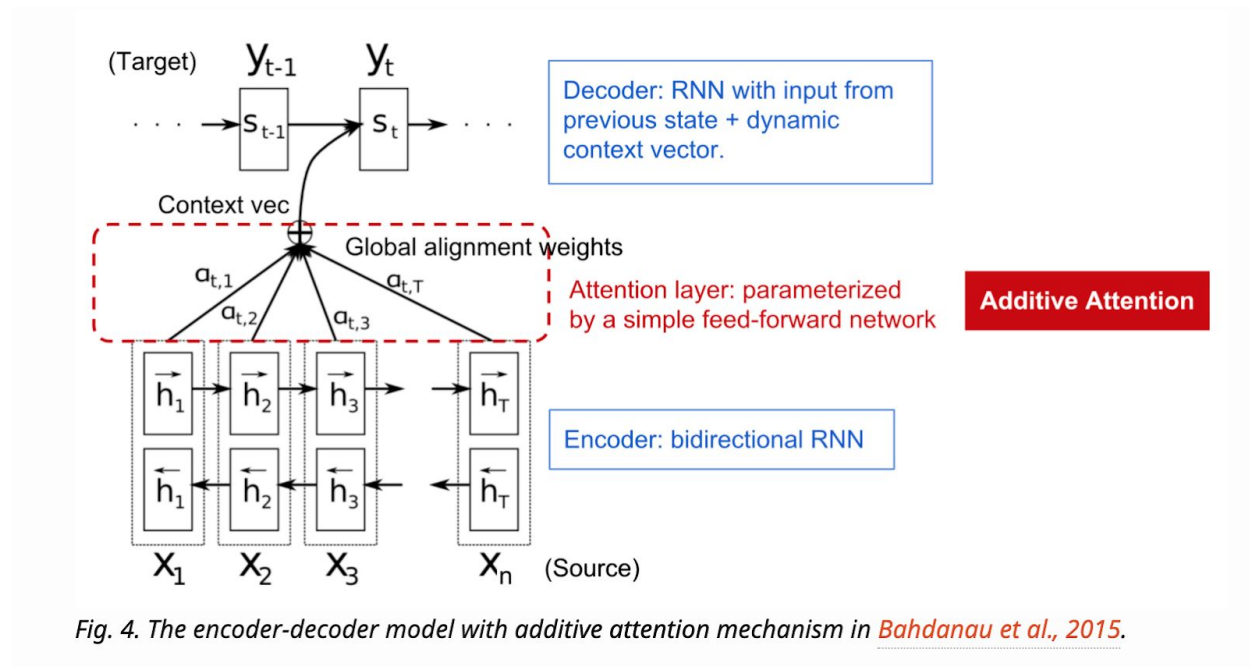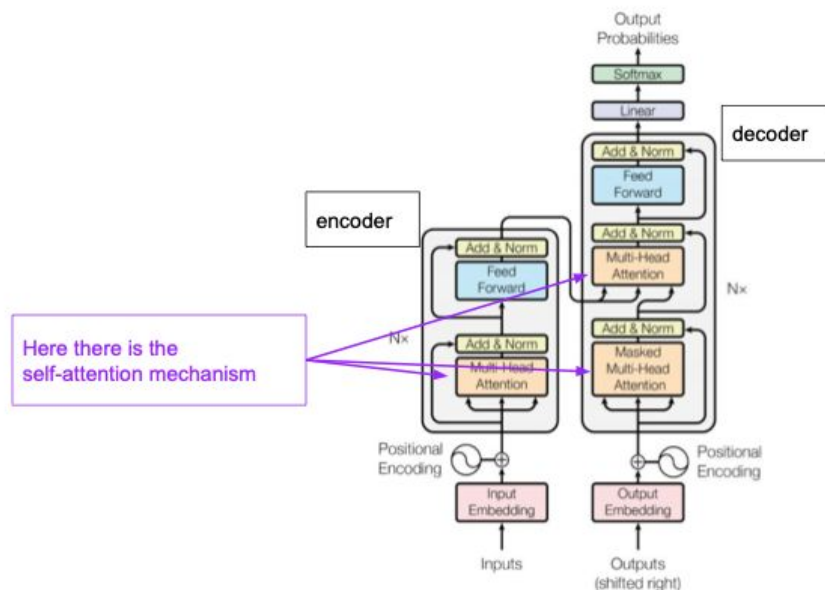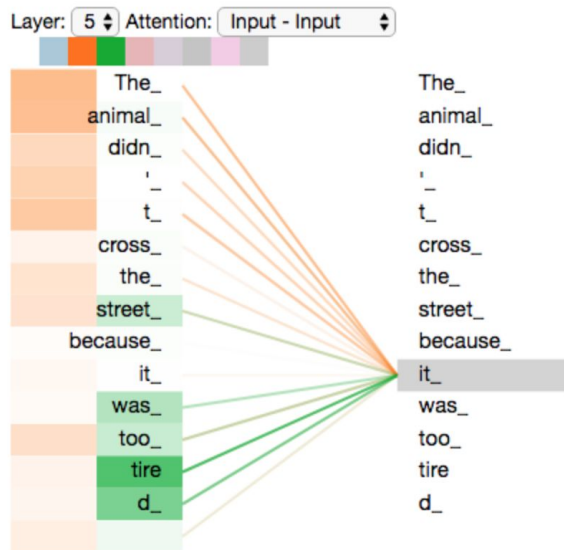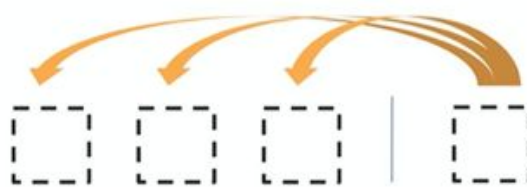

Figure 1: The Transformer - model architecture.

Slides by Alessandro Raganato (Figure 1 from Transformer paper)

Layer: 5 Attention: Input - Input

The_   The_
animal_   animal_
didn_   didn_
'_   '_
t_   t_
cross_   cross_
the_   the_
street_   street_
because_   because_
it_   it_
was_   was_
too_   too_
tire   tire
d_   d_

Three ways of attention in encoder-decoder

Encoder-Decoder Attention

Encoder Self-Attention

MaskedDecoder Self-Attention