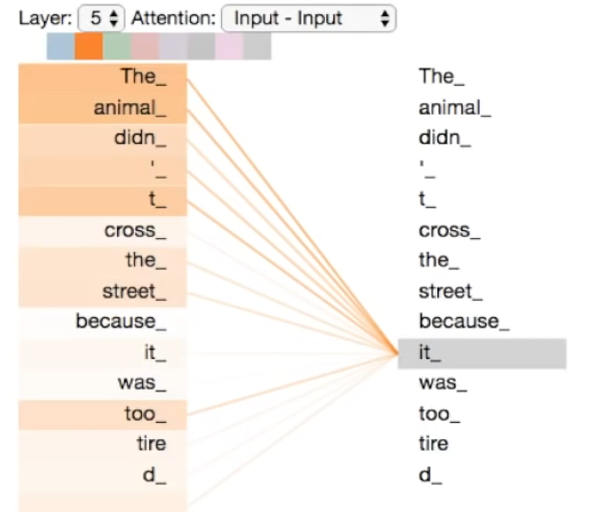
F-A.I-nal submission: Harnessing BERT.

Google has a Deep Learning NLU model called BERT. The goal of BERT is the same as ours; to build a pre-trained **language representation model**. It has three key qualities which give it the ability to meet that goal and be considered ‘High Performance’:

1. **Bidirectional**: allows for the best contextual representation of any word in a sentence.
2. **Generalisable**: Pre-trained BERT model can be fine-tuned easily for downstream NLP tasks such as Sentiment Analysis and Language Inference. Very few new parameters need to be learned.
3. **Universal**: Trained on Wikipedia and BookCorpus so no additional dataset is required.

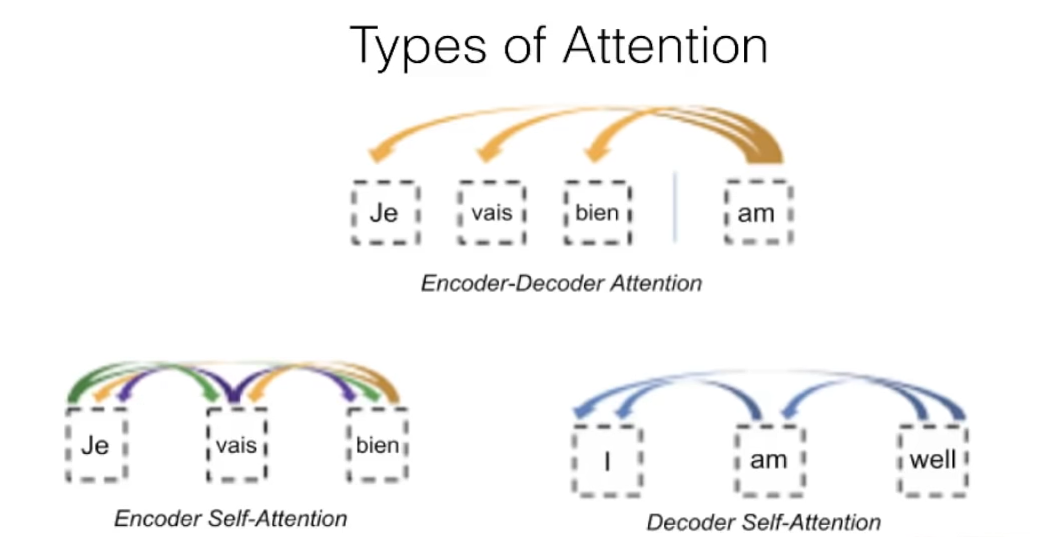
RNNs struggled with understanding what the target word of a referent is in a sentence because it was never actually trying to contextualise it. Through ‘Self-Attention’, BERT weighs each token in a sentence and is highly accurate at solving Winograd Schema examples.

Consider the following visualisation:



As humans, we intuitively know that the subject of the referent is the animal however, a computer cannot infer this so easily. Note that the strength of the link between **animal\_** and **it\_** is much darker than the link between **street\_** and **it\_.**

BERT can also be applied to Machine Translation:



RNNs use a Left-to-Right context which shows the machine the first part of a sentence and asks it to generate the next word. You can do this recursively by asking it to continuously generate new words by showing it the sentence + the word it just generated. This is how Natural Language Generation works. TensorFlow is often used for this.

By using a Bidirectional context, you are changing the principle of NLG from ‘What is the next word?’ to ‘Guess the missing word.’ which is likely to yield more realistic results because the machine knows the full context of the sentence apart from the one missing word in the middle of the sentence.