THE NARRATED TRANSFORMER LANGUAGE MODEL

this guy's explanation was the easiest to understand

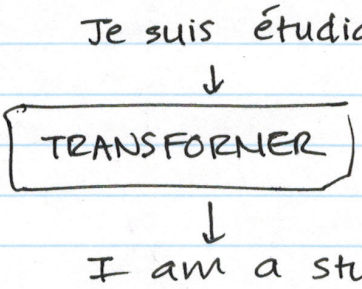
The transformer is the tech that underlies a lot of the progress in the last few years of NLP. (paper came out of Google & U of T in 2017). A transformer is a type of Neural network, just as CNN & RNNs are. It's a specific architecture (layout) of the NN.

BERT & its derivatives are popular transformer-based NNs right now.

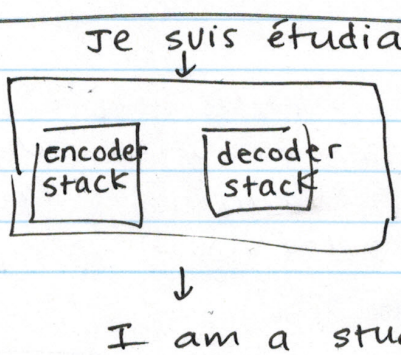
Transformers are not limited to NLP, but also have applications in computer vision, music generation, ... (Byebye CNN — transformers are gaining recognition!)

ARCHITECHTURE

OVERVIEW



DETAILED VIEW



TRAINING

get random large amount of text → [train model] → trained model

if the model gets it wrong, we tell it what the answer SHOULD HAVE BEEN & calculate an error.

COMPONENTS

FFNN (Feed-forward NN) → project output, P(tokens) ← self-attention ← token embeddings (matrix) ← tokenization

which gives context to tokens