Introduction to Machine Learning

All for One

Choi Minjoo

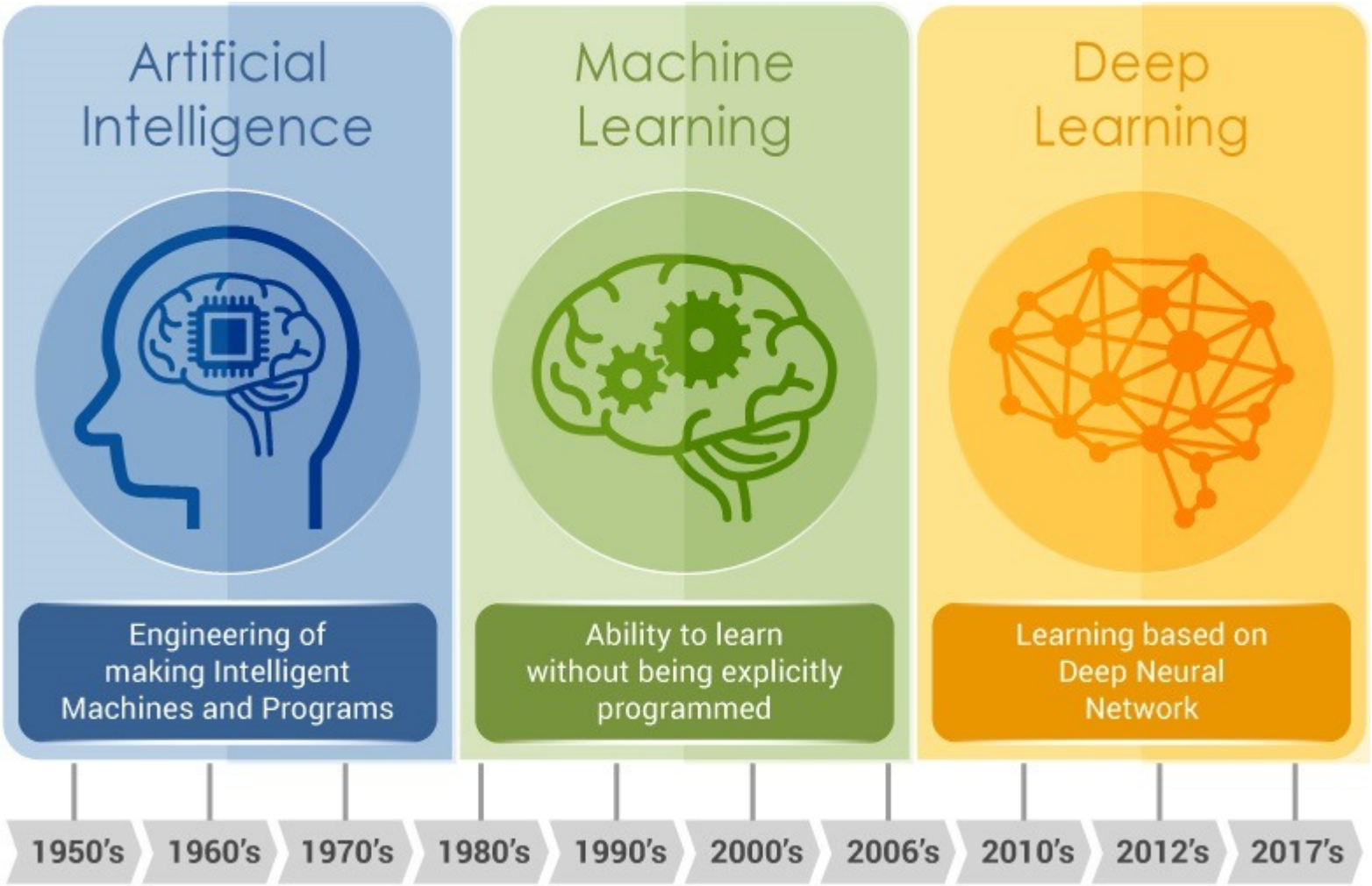
Choi Minjoo (Judy Choi)

- Careers
  - 2012 ~ 2019 **Malware Analyst**
  - 2018 **Working Holiday** in France
  - 2020 ~ **M.S** in Kangwon Univ
    - Intelligence Software Lab
      - **NLP (Machine Translation)**
  - 2021 ~ 2022 **Bering Lab** (NLP Researcher & Engineer)
- SNS
  - https://www.facebook.com/minjoo.choi.562/
  - https://github.com/Judy-Choi

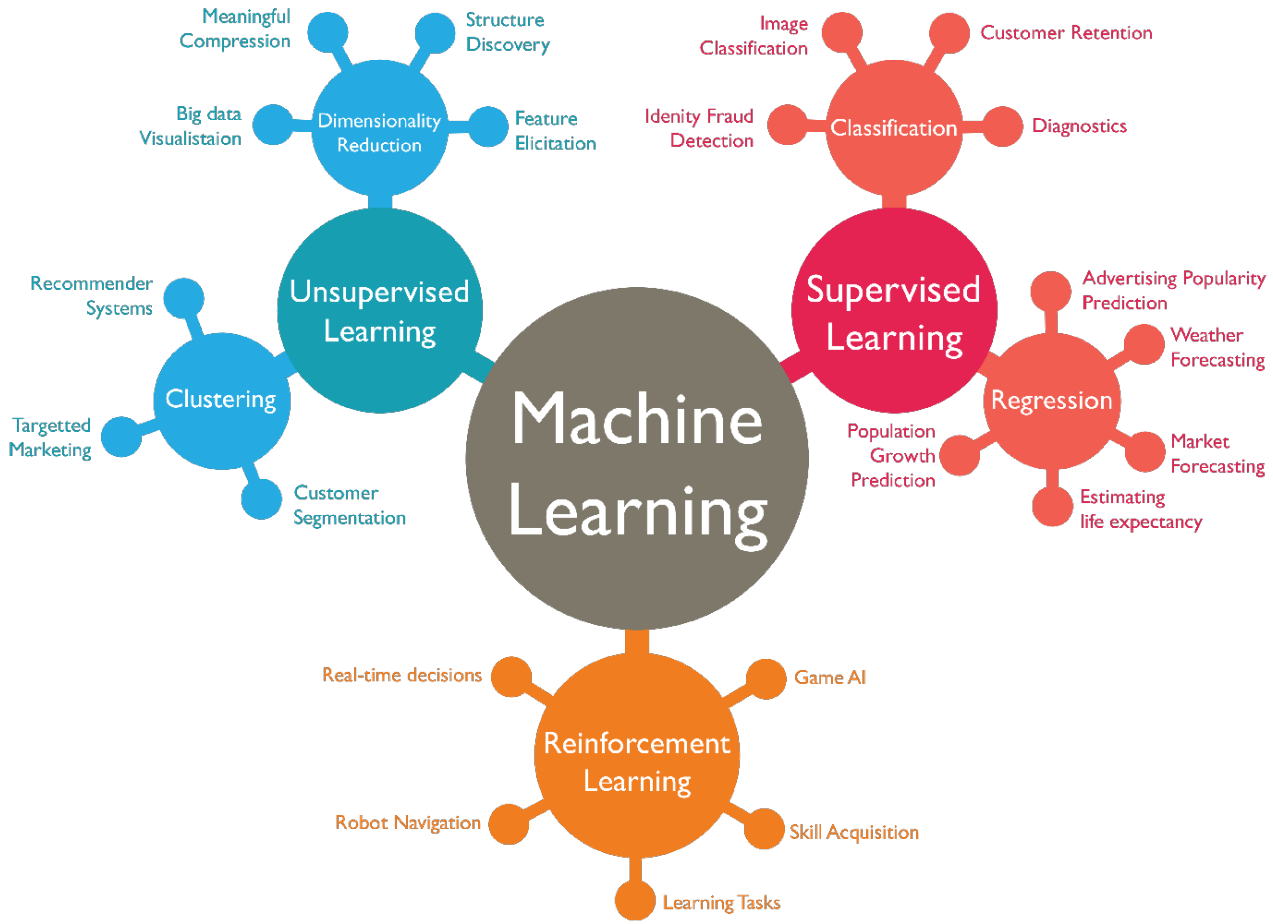
Contents

Abstract of Deep Learning

from A.I to Deep Learning...

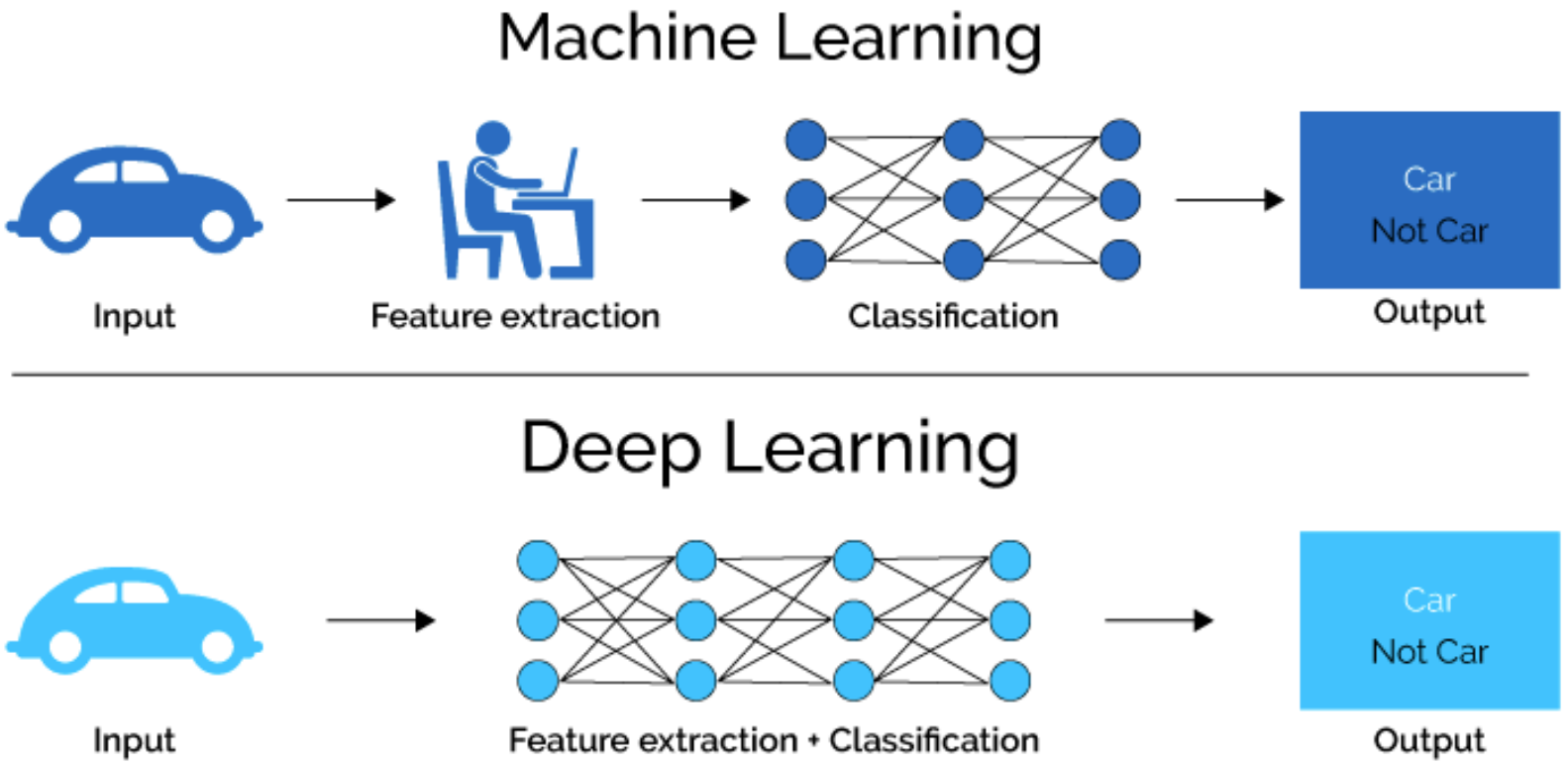


Machine Learning



https://medium.com/intro-to-artificial-intelligence/reinforce-a-policy-gradient-based-reinforcement-learning-algorithm-84bde440c816

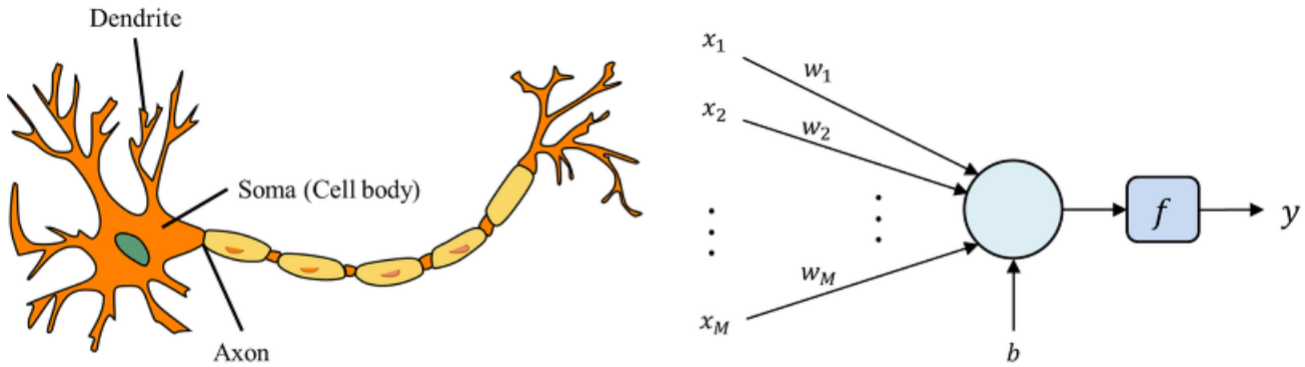
Machine Learning vs Deep Learning



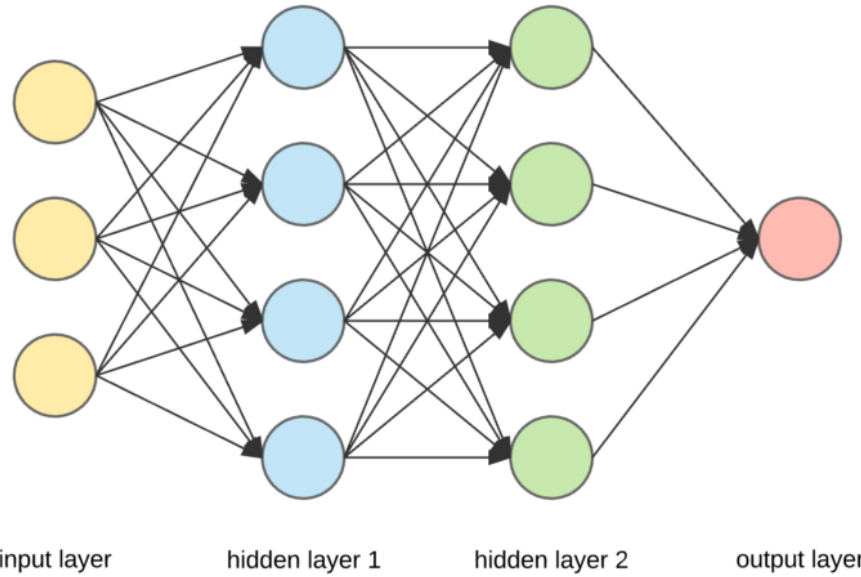
https://lawtomated.com/a-i-technical-machine-vs-deep-learning/

Deep Learning

- Neural Network
  - Artificial neural network mimicking neural networks in the human brain.
  - Brain : Neuron = A.I : Perceptron



[그림 1] 생물체의 neuron (좌)과 artificial neuron (우)



https://smartstuartkim.wordpress.com/2019/01/27/history-of-neural-networks-1-perceptron/

https://gongster.medium.com/how-does-a-neural-network-work-intuitively-in-code-f51f7b2c1e3f

Deep Learning

- Deep Neural Network (DNN)
  - A network of multiple layers of perceptron layers

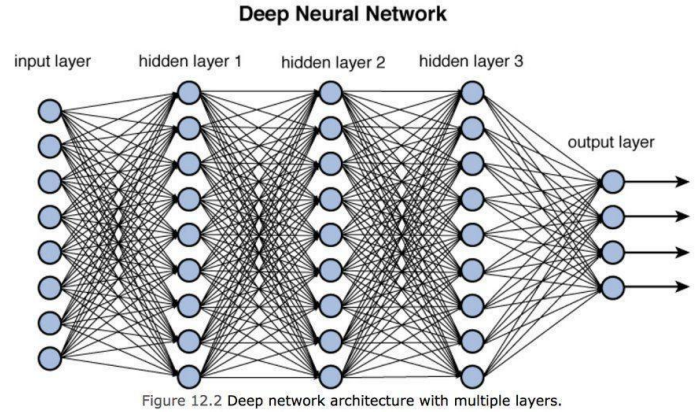
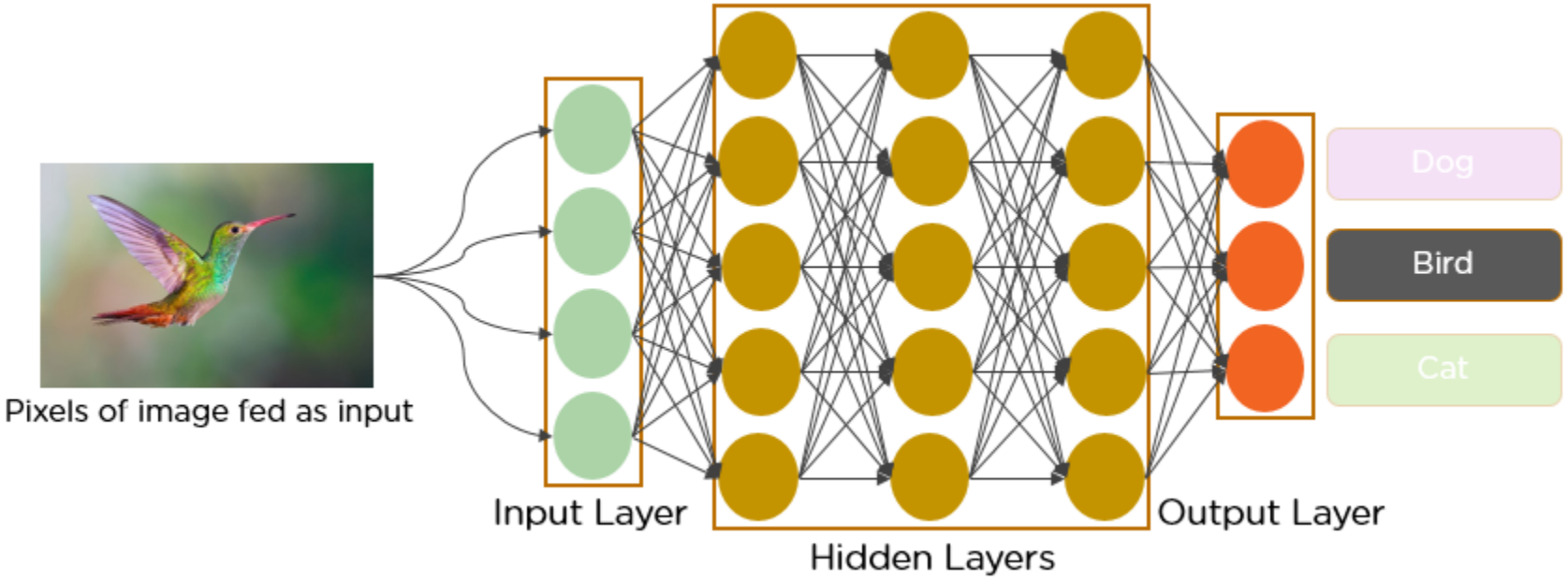


Figure 12.2 Deep network architecture with multiple layers.

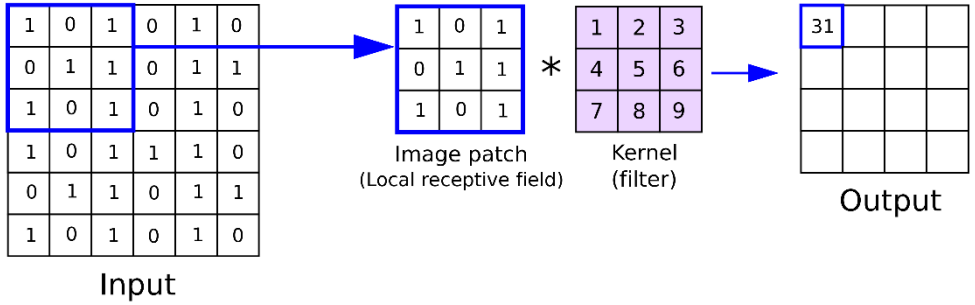


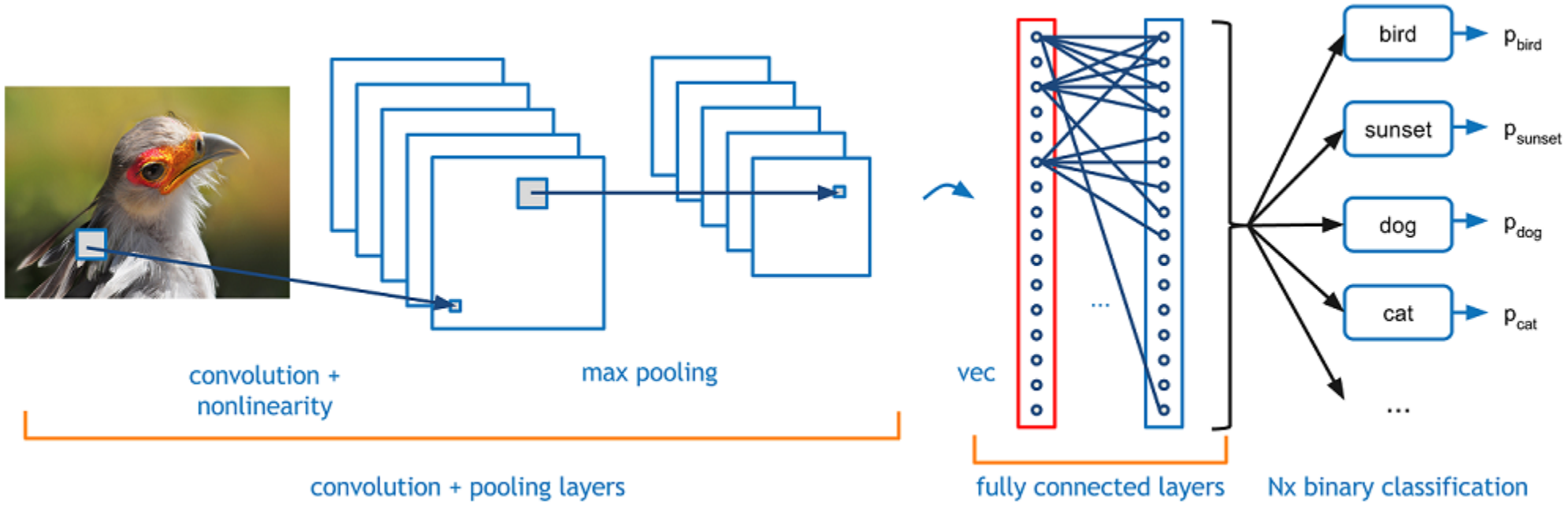
https://towardsdatascience.com/training-deep-neural-networks-9fdb1964b964

https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

Deep Learning

- CNN (Convolution Neural Network)

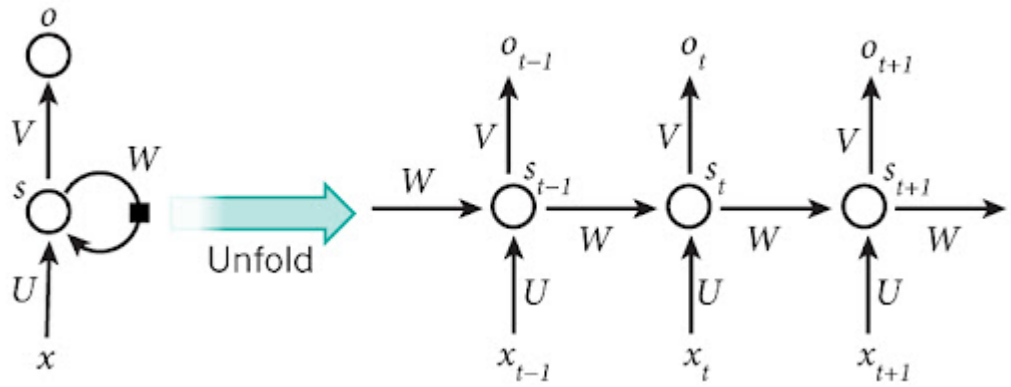


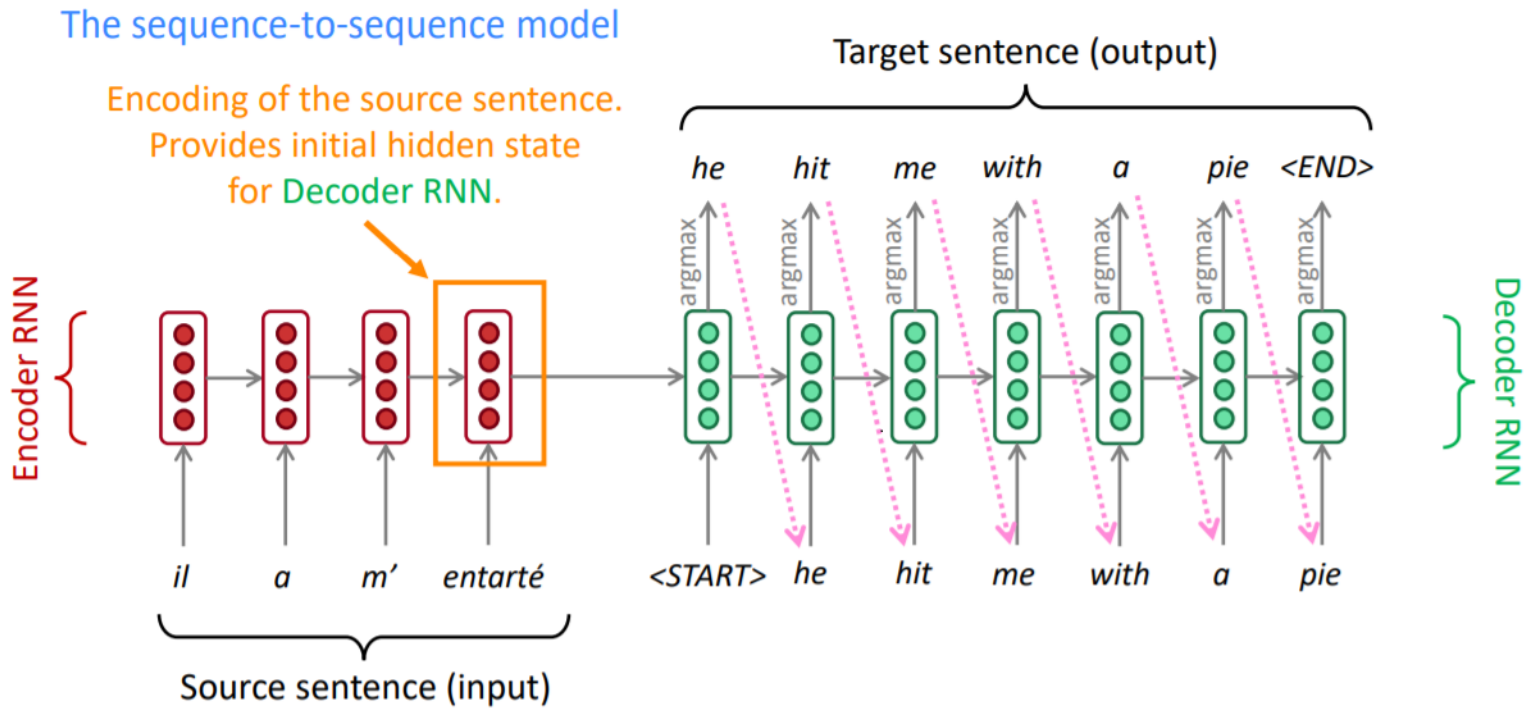


https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529

Deep Learning

- RNN (Recursive Neural Network)





https://aikorea.org/blog/rnn-tutorial-1/

https://medium.com/analytics-vidhya/neural-machine-translation-using-bahdanau-attention-mechanism-d496c9be30c3

Natural Language Processing (NLP)

NLP?

- ‘Interactions between computers and human language’
- A subfield of linguistics, computer science, and artificial intelligence
- Programming computers to process and analyze large amounts of natural language data.
- The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them.

Field of application

- Chatbot
  - Alexa, Siri, Bixby.
- Text Summarization
  - Google search
- Document Classification
  - Spam mail filtering
- Machine Translation
  - Google Translate

Uzbekistan

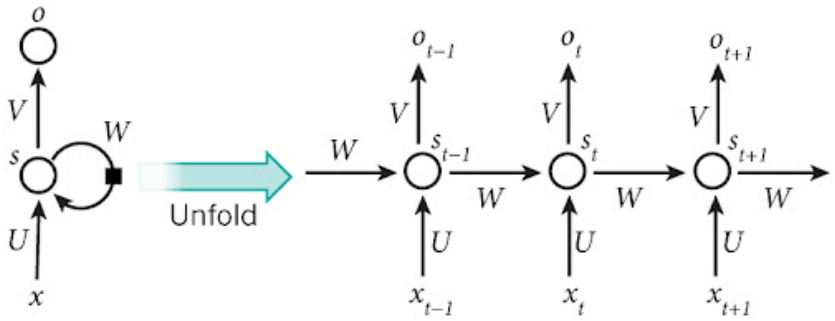
Country in Central Asia

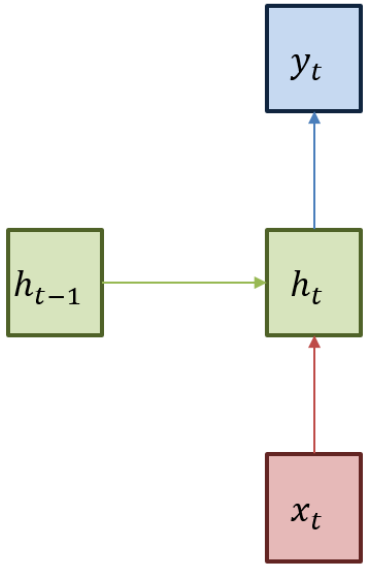
Uzbekistan is a Central Asian nation and former Soviet republic. It's known for its mosques, mausoleums and other sites linked to the Silk Road, the ancient trade route between China and the Mediterranean. Samarkand, a major city on the route, contains a landmark of Islamic architecture: the Registan, a plaza bordered by 3 ornate, mosaic-covered religious schools dating to the 15th and 17th centuries. — Google

https://www.mvtvwireless.com/why-are-my-emails-going-into-spam/

Google Translate

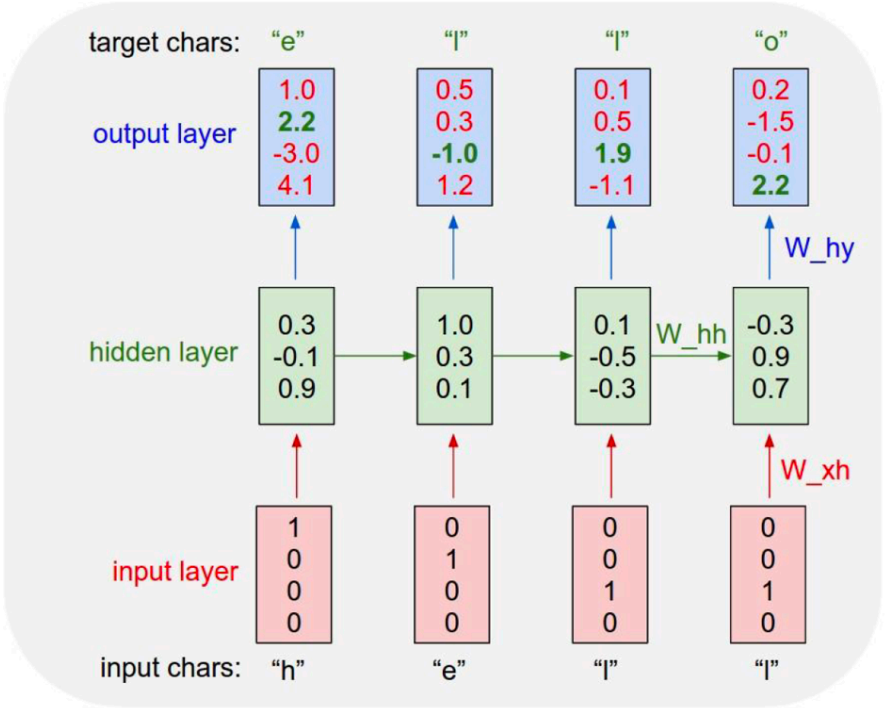
RNN





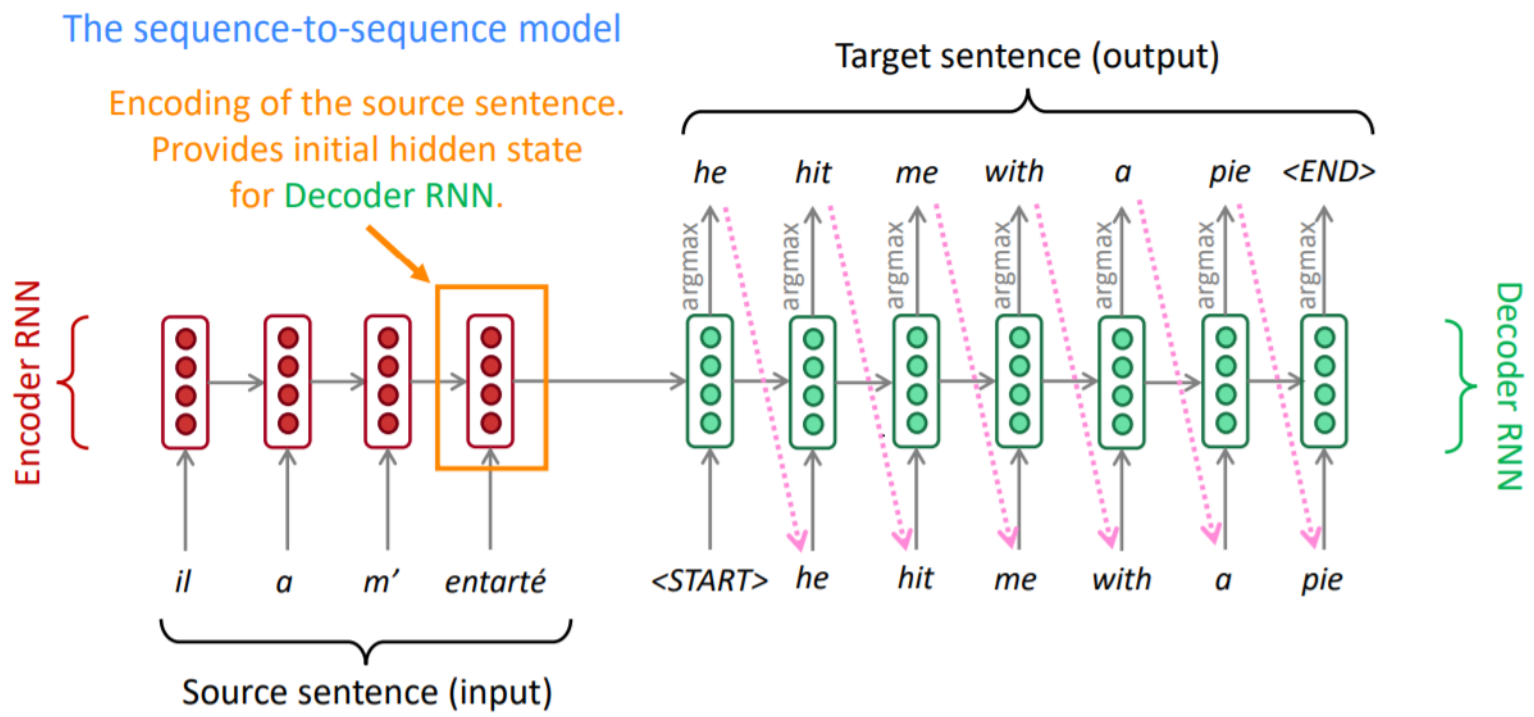
$$y_t = W_{hy} h_t + b_y$$

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t + b_h)$$



https://aikorea.org/blog/rnn-tutorial-1/
https://ratsgo.github.io/natural%20language%20processing/2017/03/09/rnnlstm/

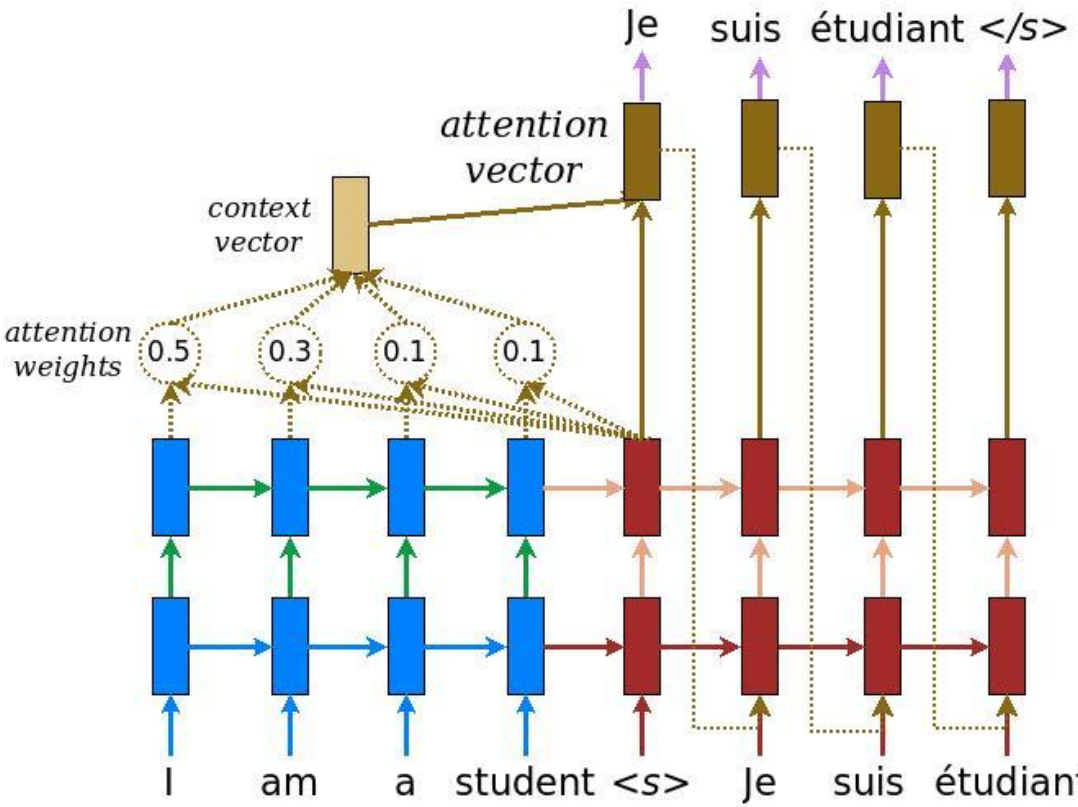
Seq2Seq



https://medium.com/analytics-vidhya/neural-machine-translation-using-bahdanau-attention-mechanism-d496c9be30c3

Seq2Seq + '**Attention**'

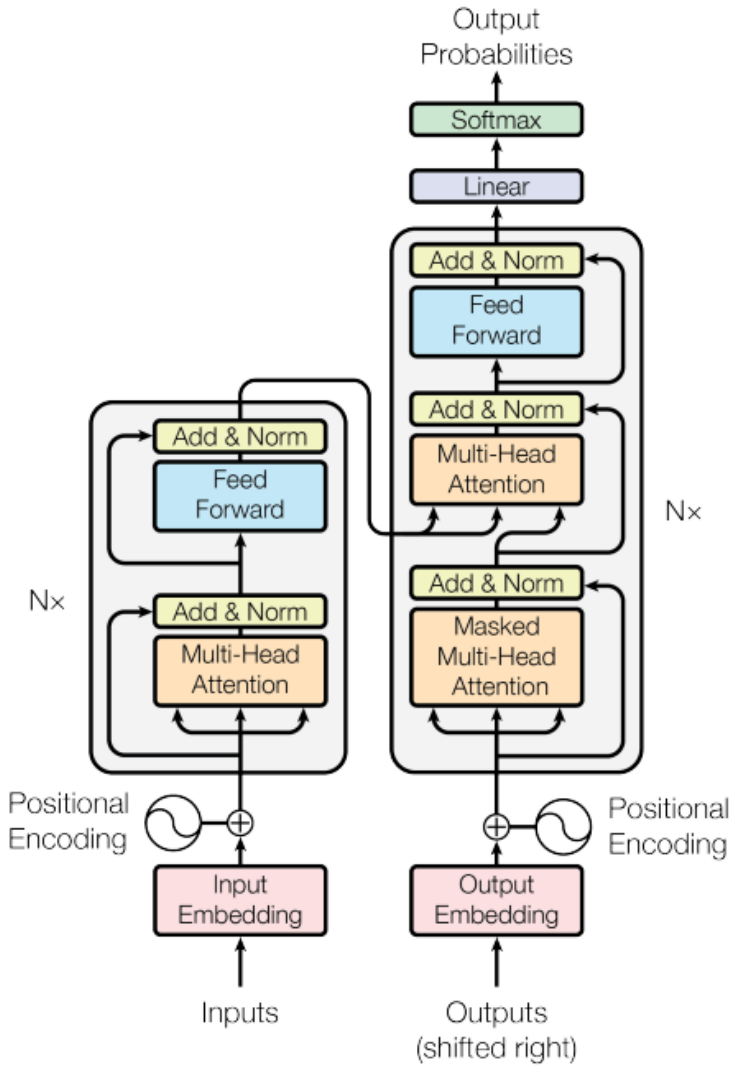
- Decoder learn to **focus** over a specific range of the input sequence



https://www.tensorflow.org/text/tutorials/nmt_with_attention

Transformer

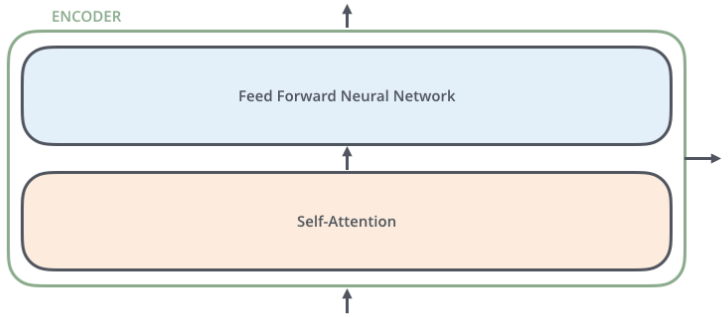
- ‘Attention is All you Need’

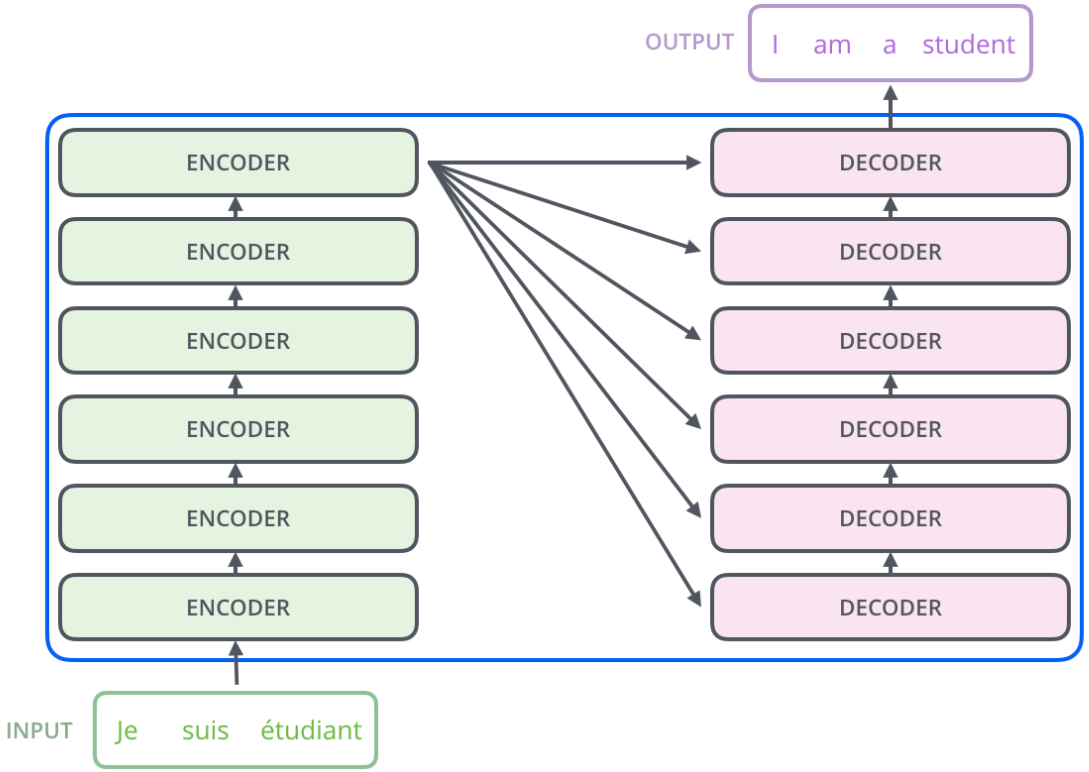


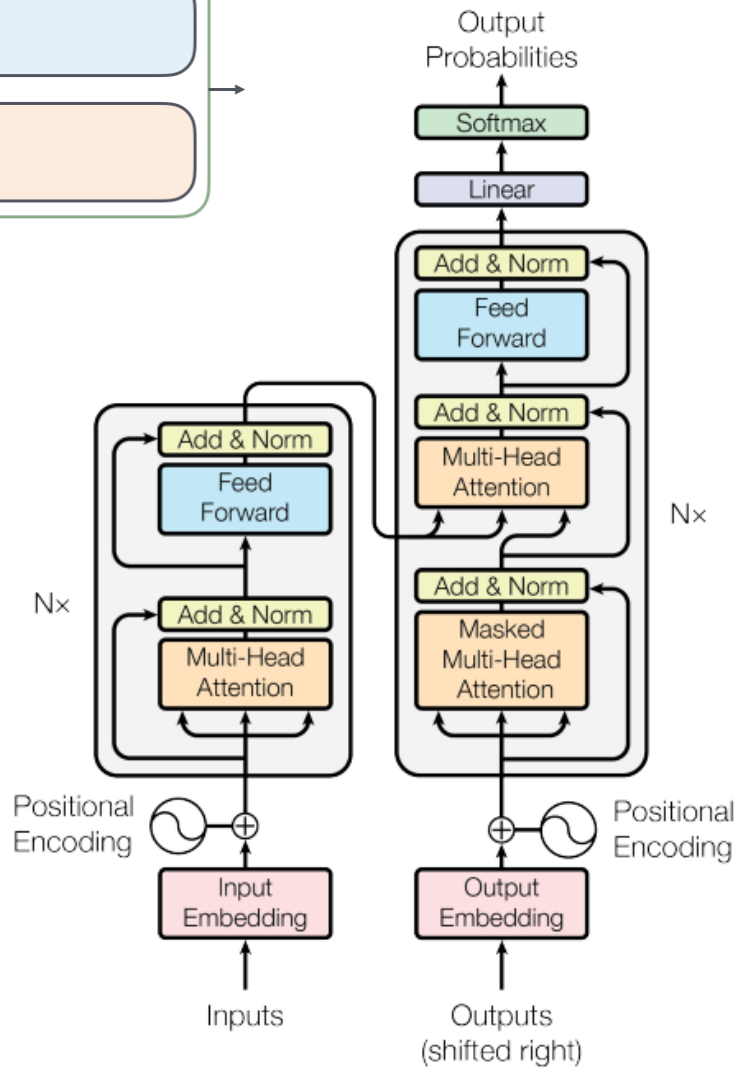
https://jalammar.github.io/illustrated-transformer/

Transformer

- Encoder – Decoder (6-6)



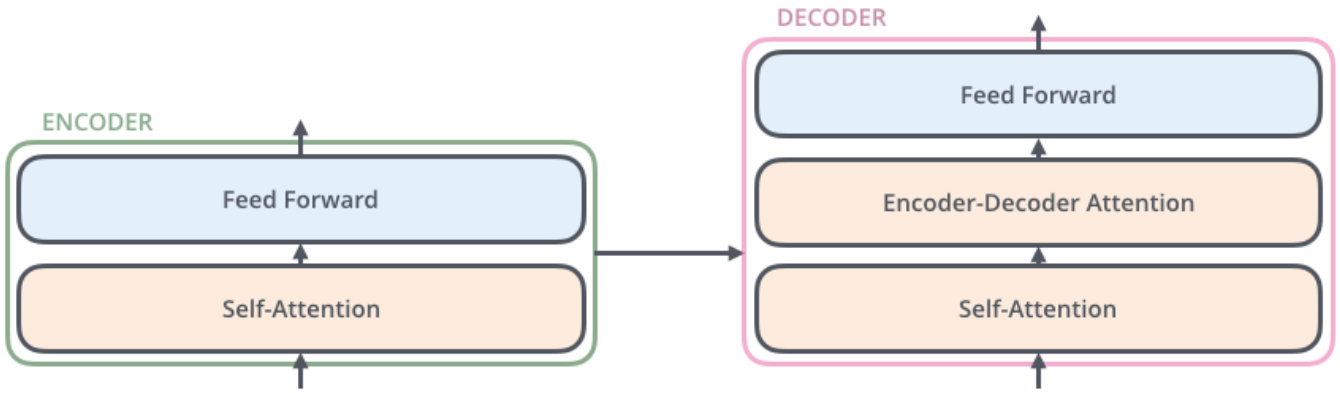


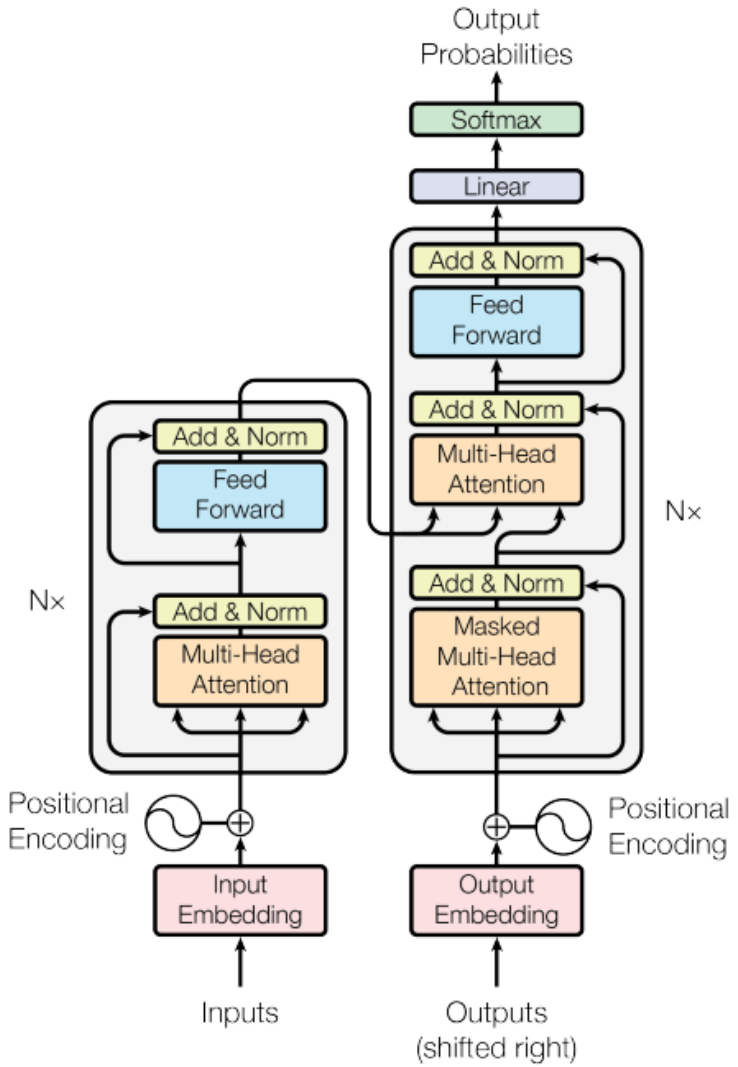


https://jalammar.github.io/illustrated-transformer/

Transformer

- FFNN & Self-Attn in Encoder

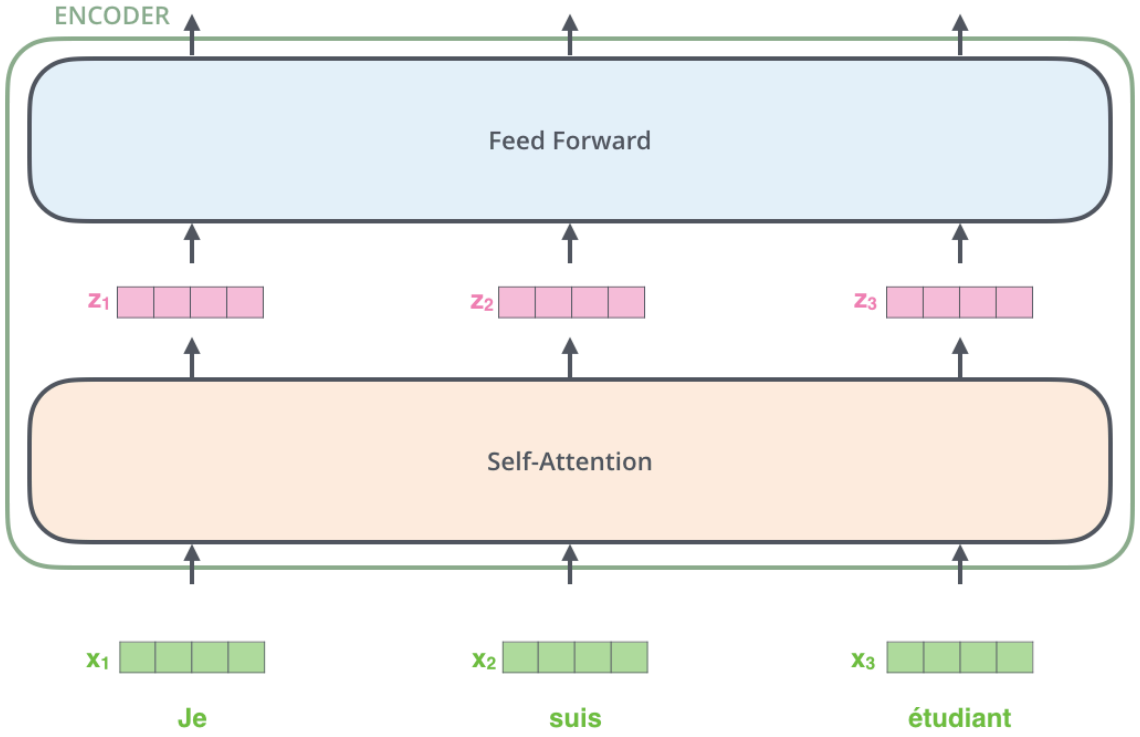


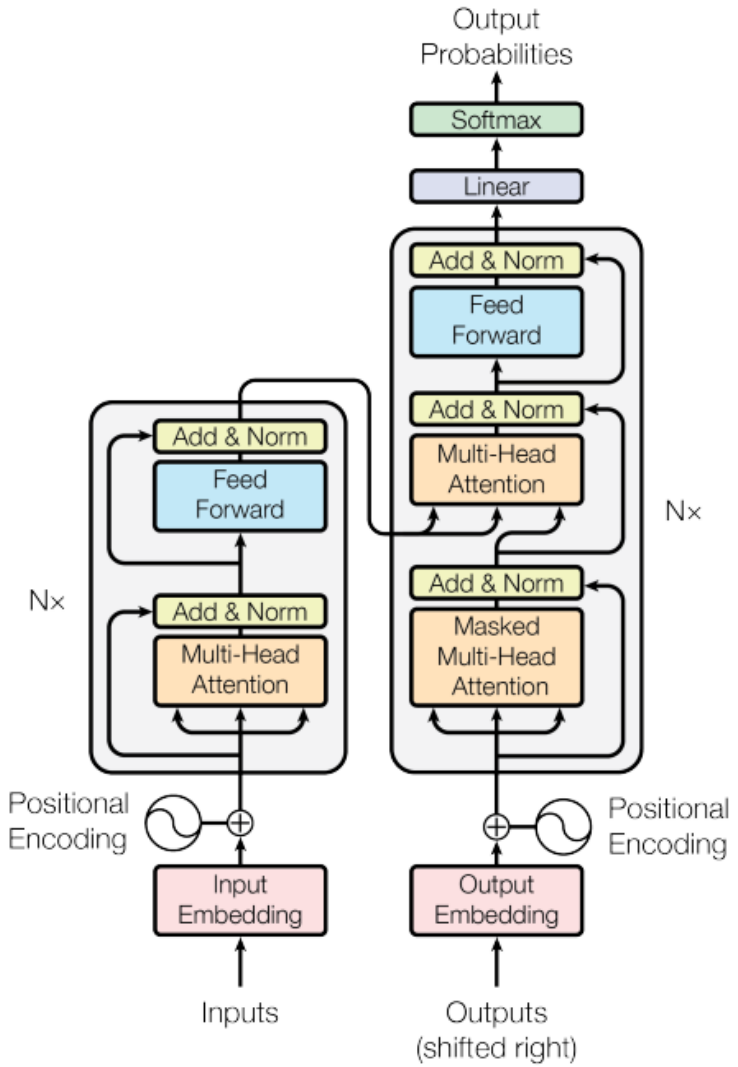


https://jalammar.github.io/illustrated-transformer/

Transformer

- Input embedding

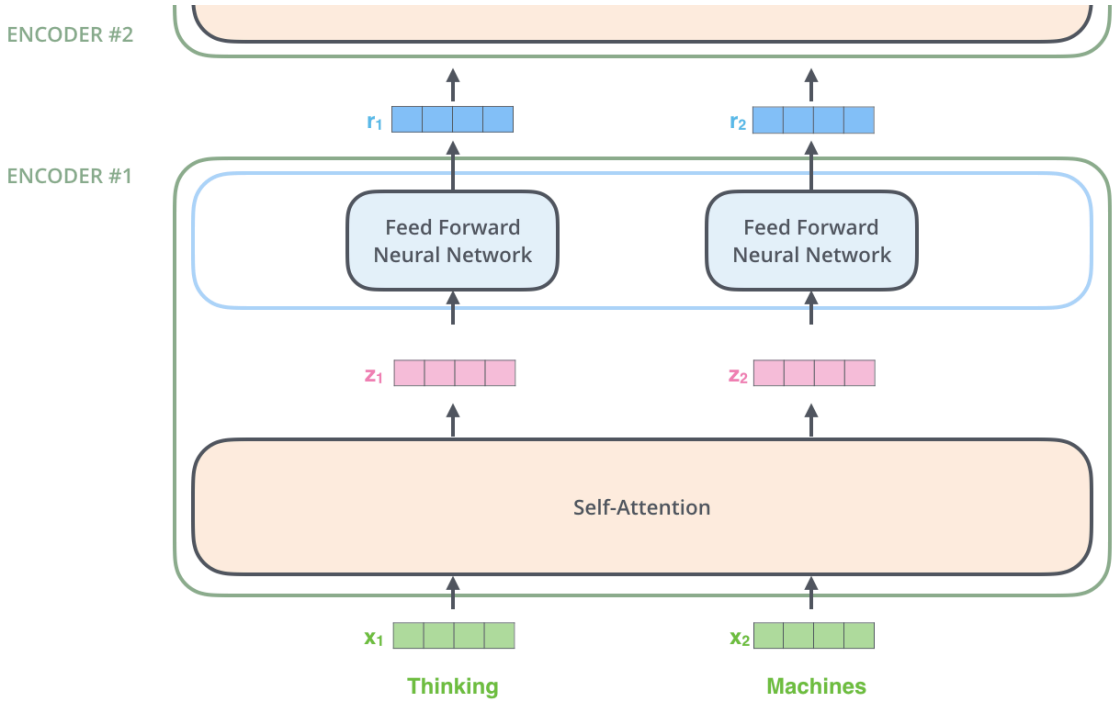


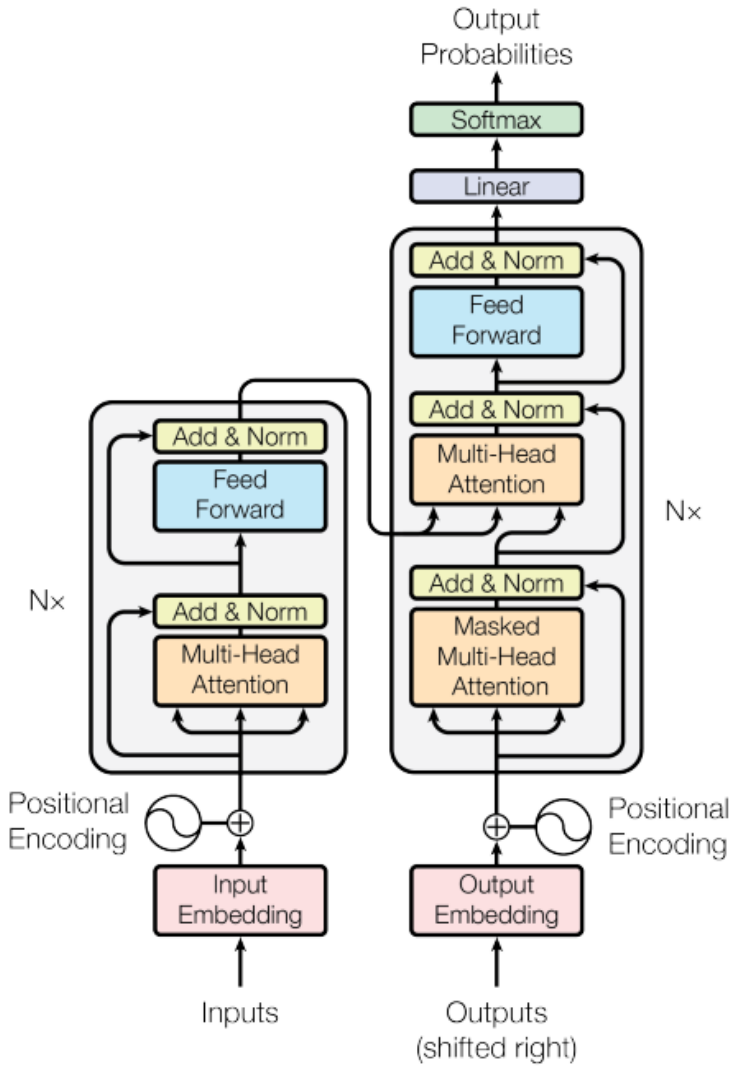


https://jalammar.github.io/illustrated-transformer/

Transformer

- Encoding

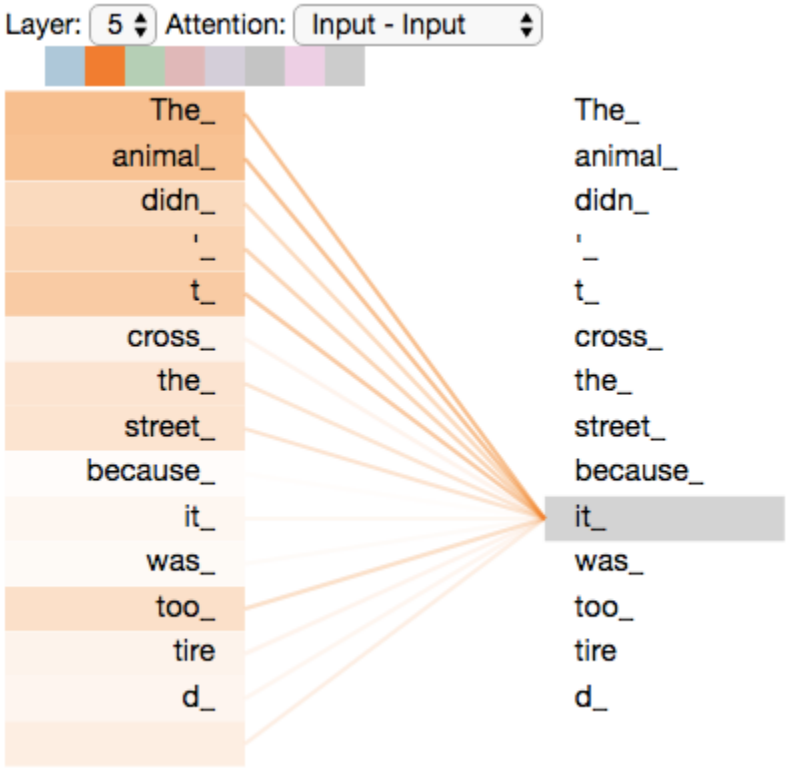


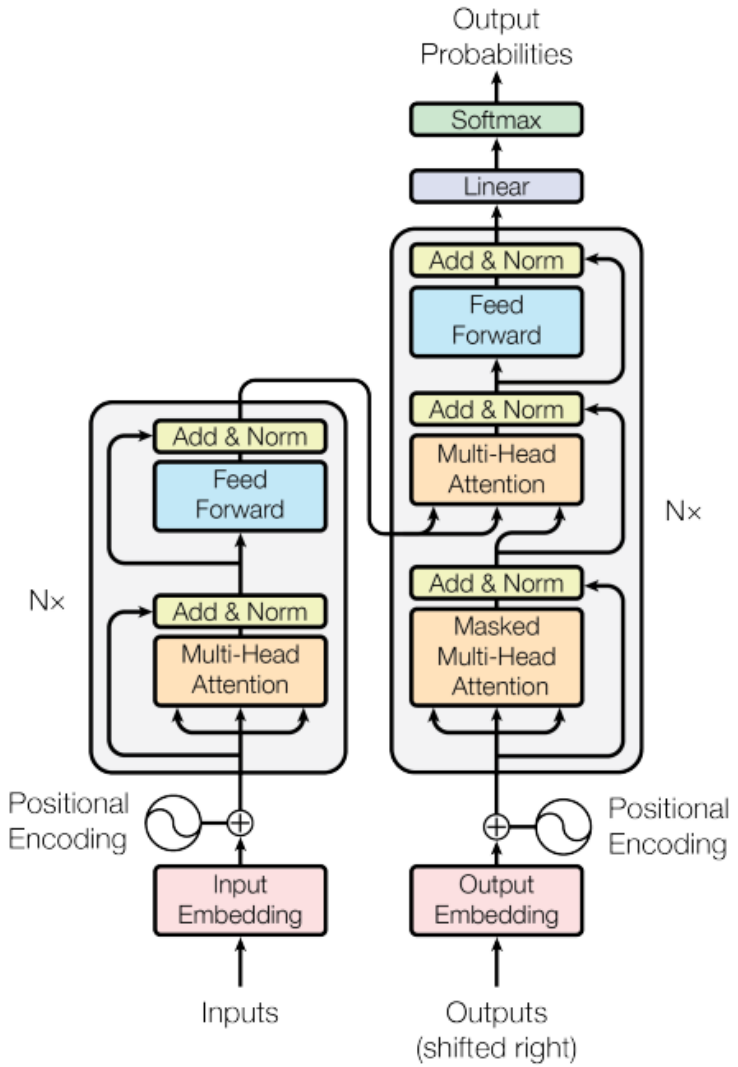


https://jalammar.github.io/illustrated-transformer/

Transformer

- Self-Attention

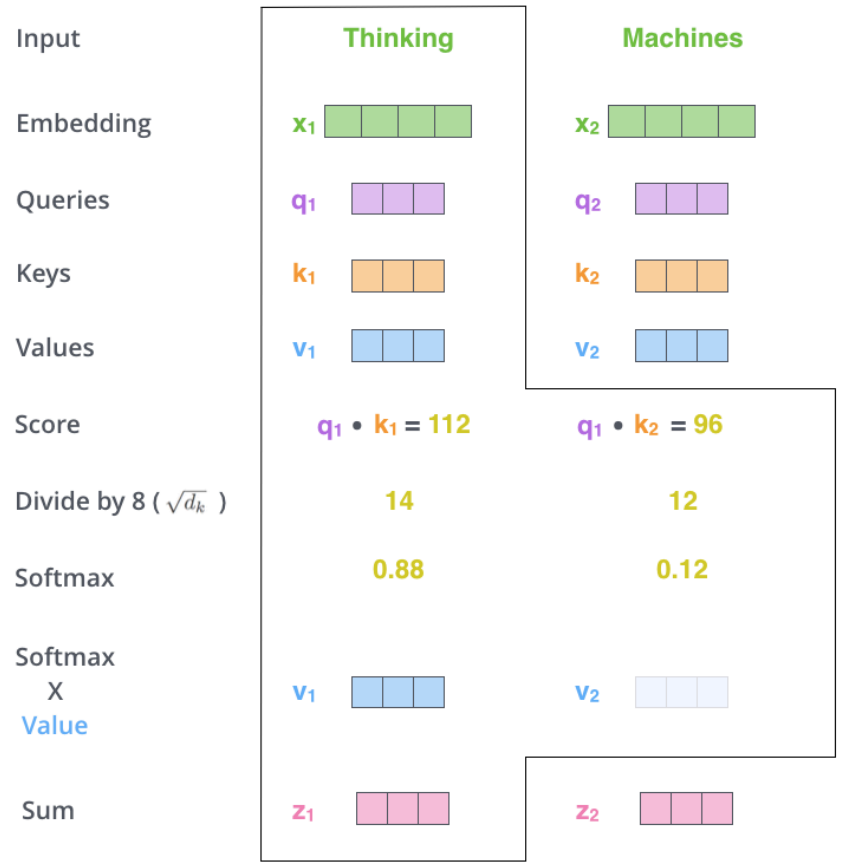


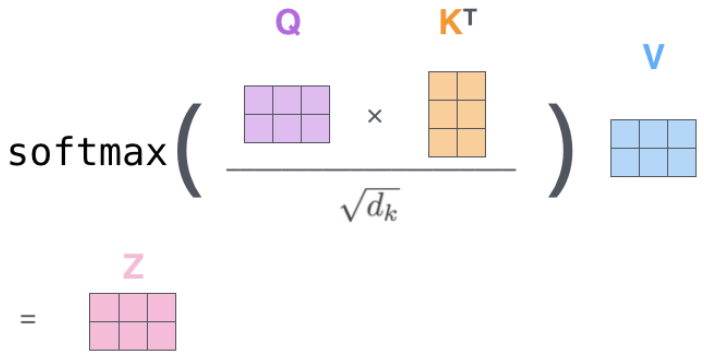


https://jalammar.github.io/illustrated-transformer/

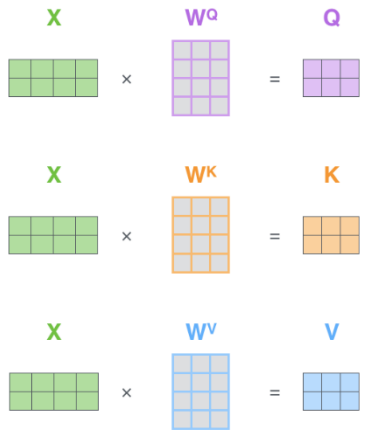
Transformer

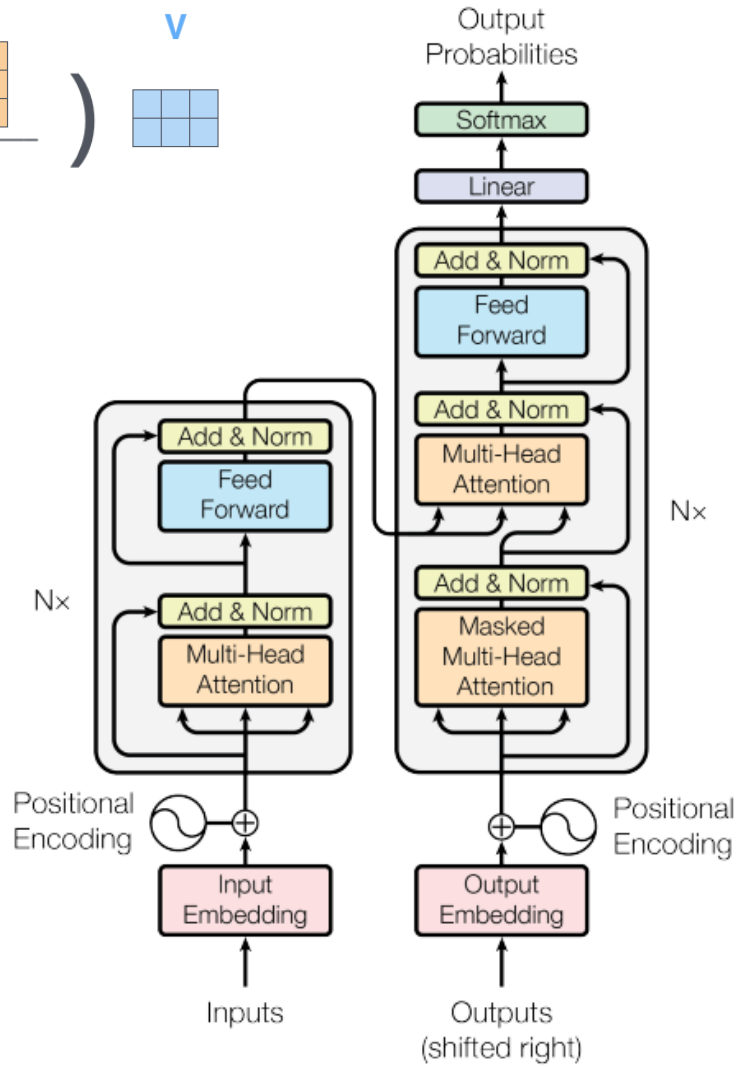
- Self-Attention





Matrix Calculation of Self-Attention

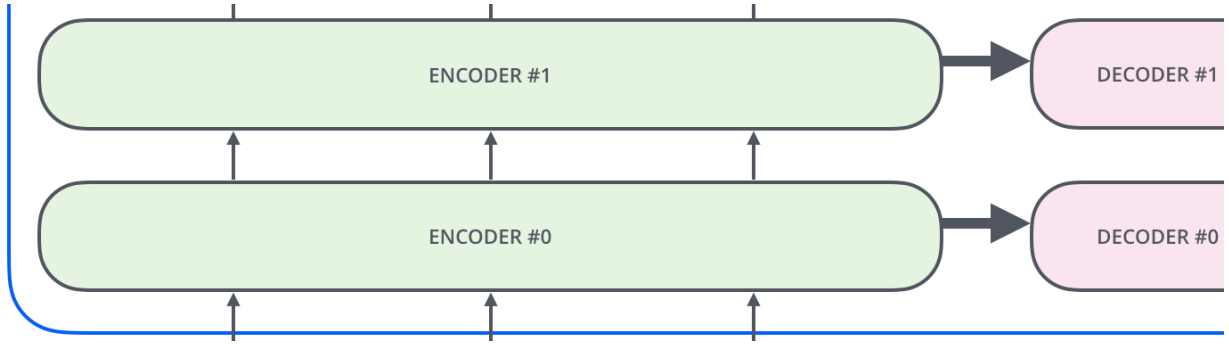


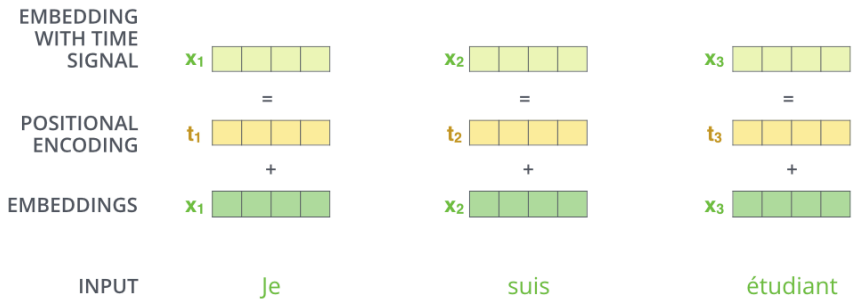


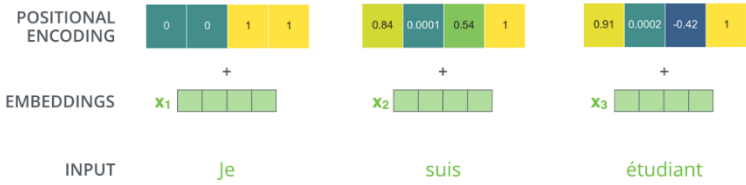
https://jalammar.github.io/illustrated-transformer/

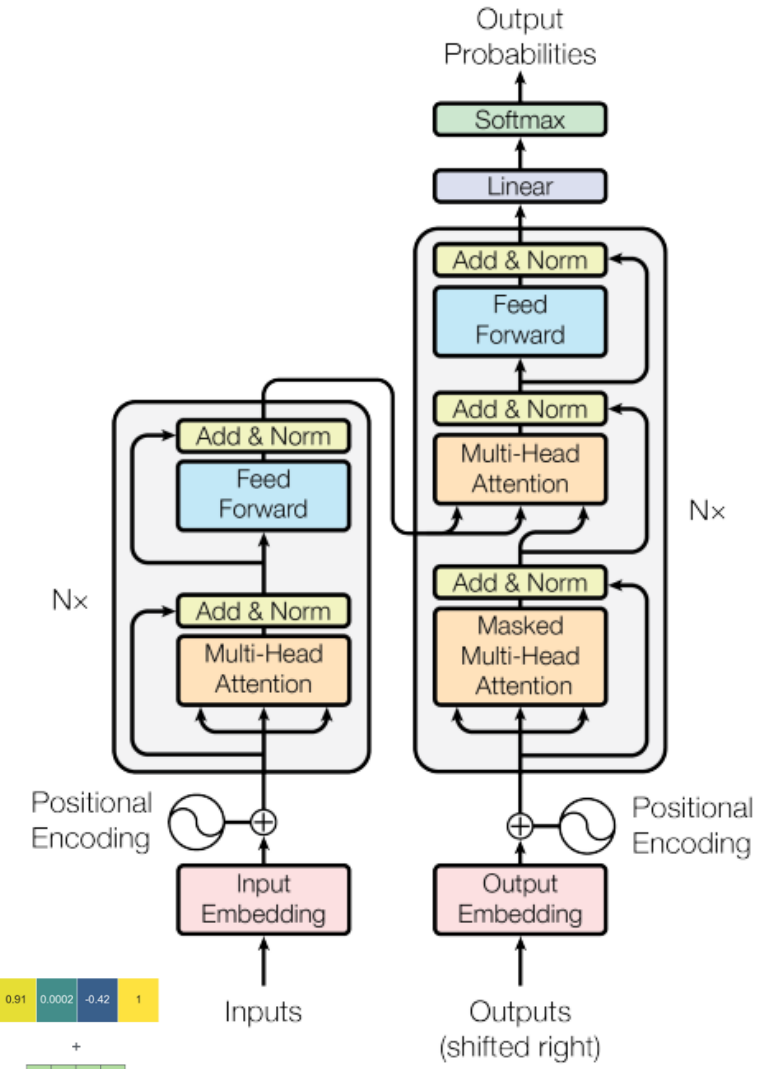
Transformer

- Positional Encoding





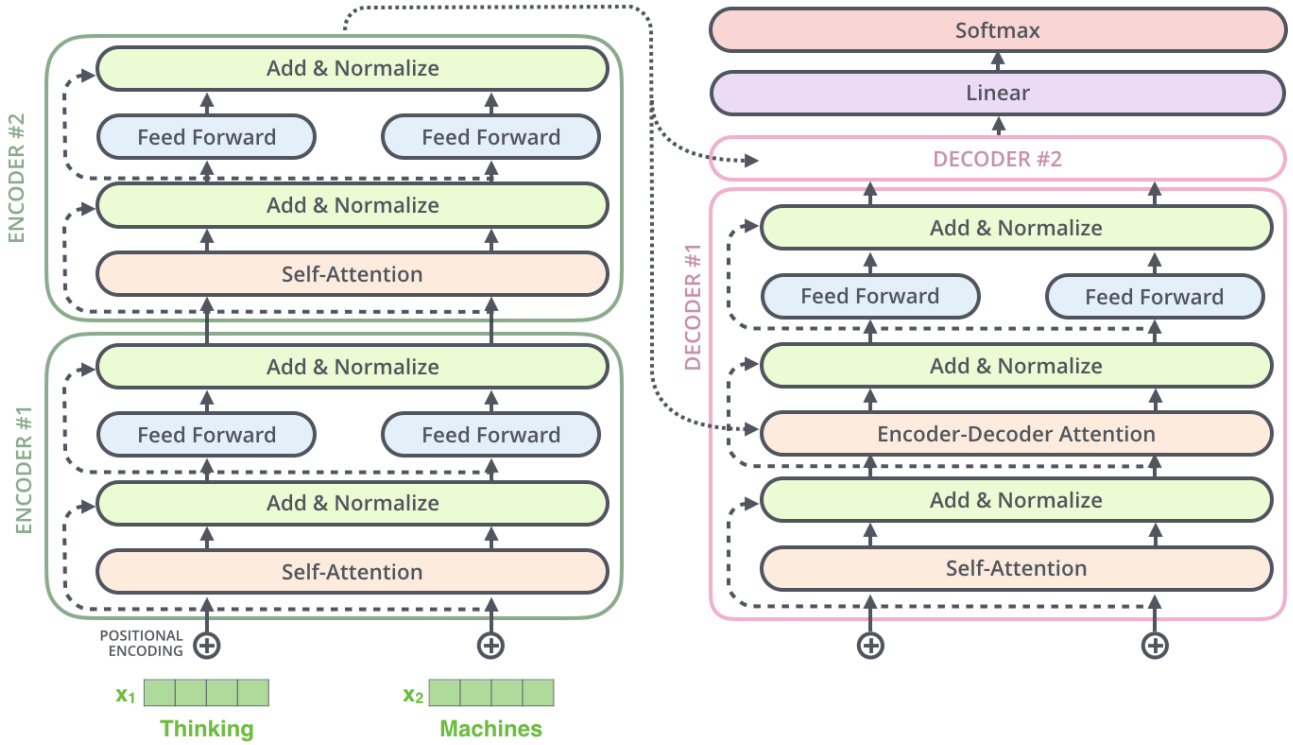


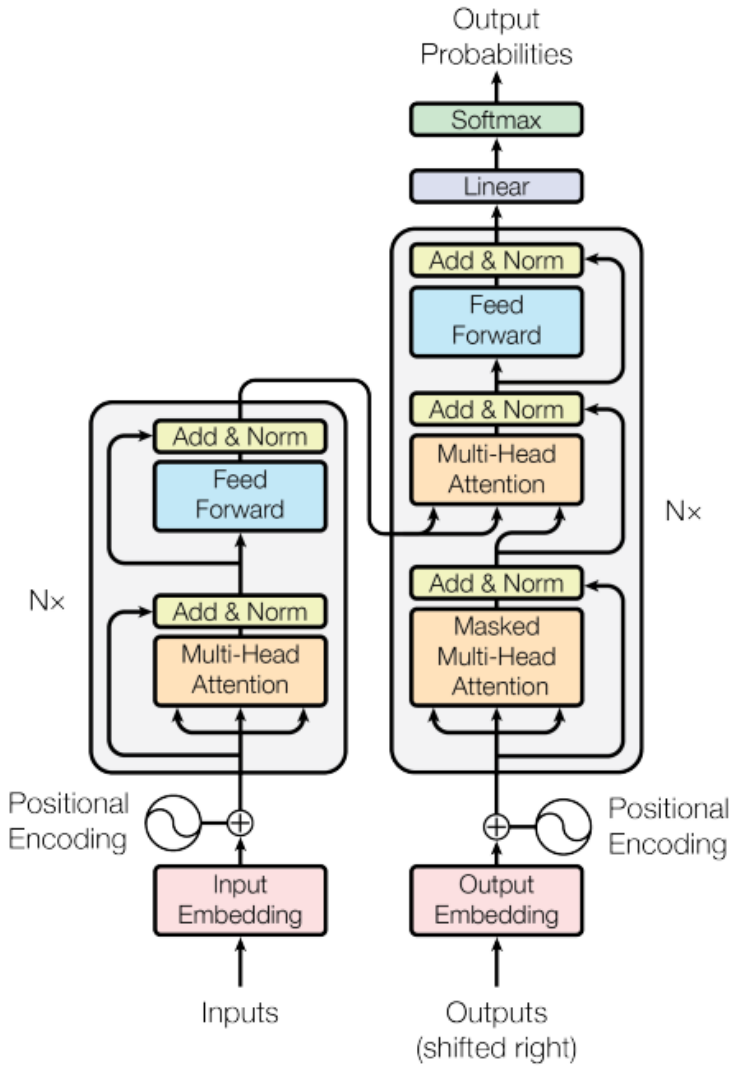


https://jalammar.github.io/illustrated-transformer/

Transformer

- Detail process in encoder & decoder

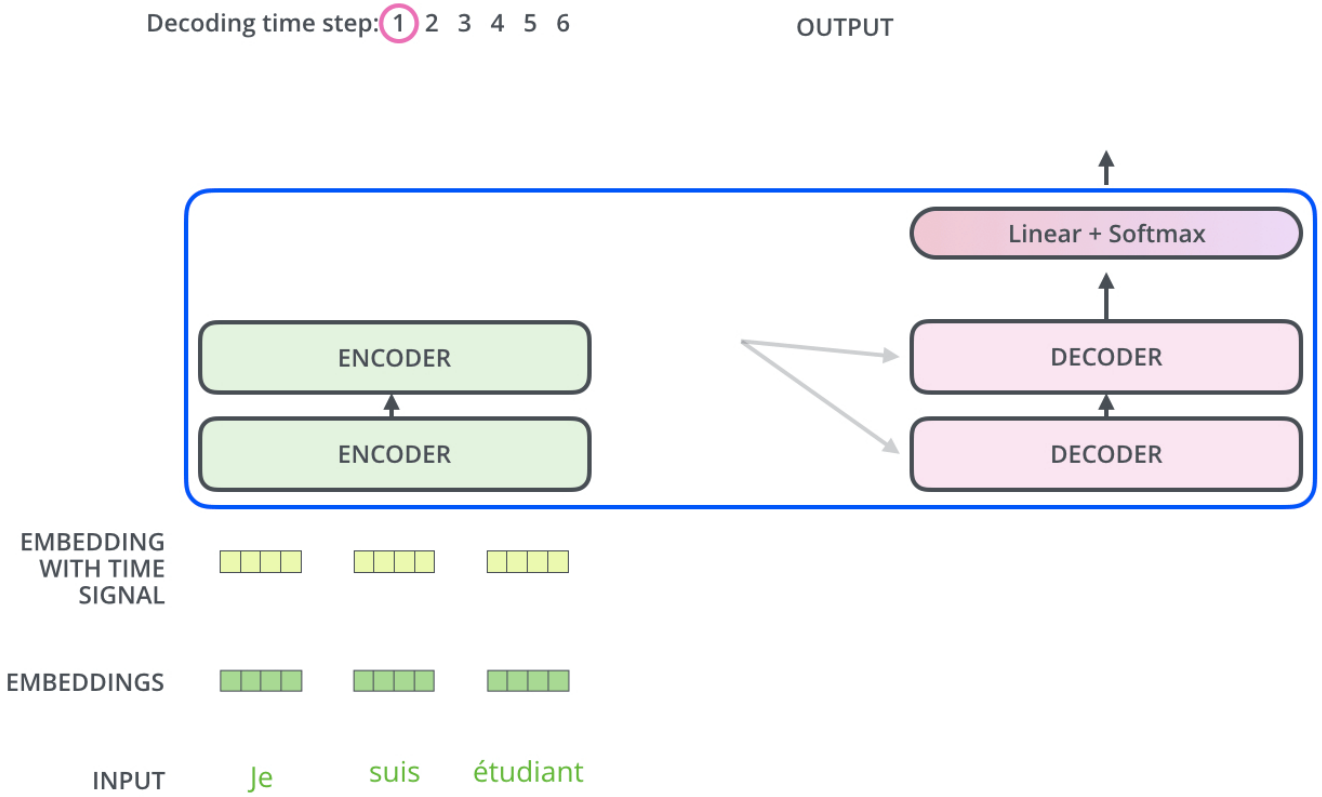


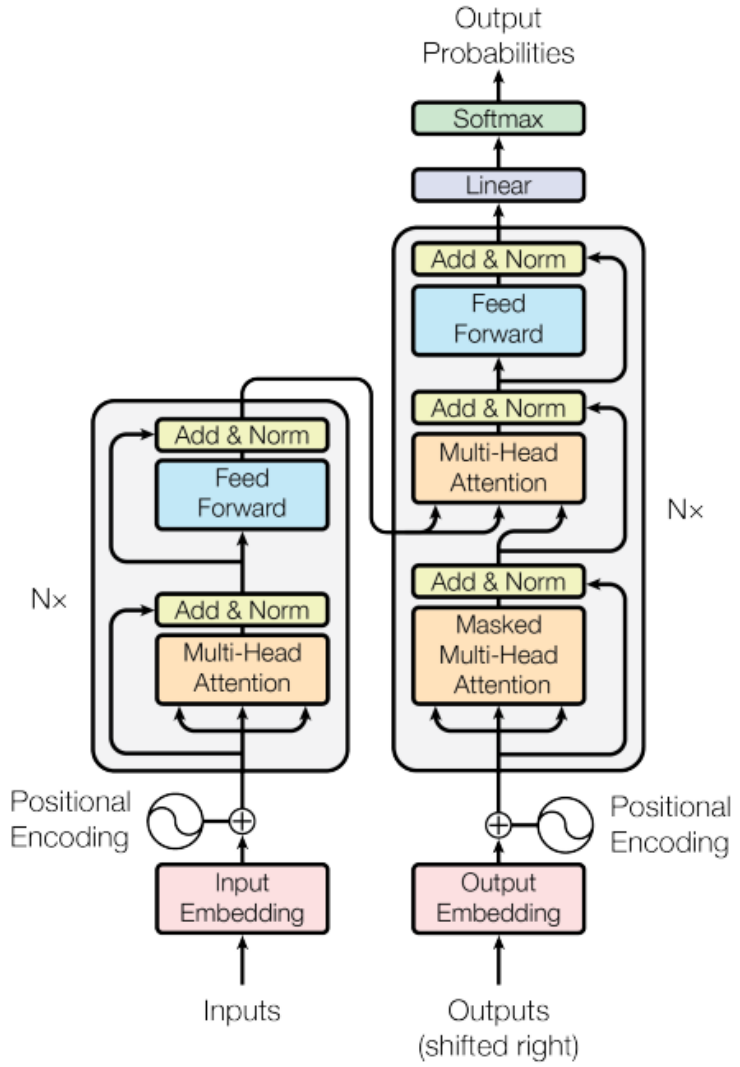


https://jalammar.github.io/illustrated-transformer/

Transformer

- Decoder

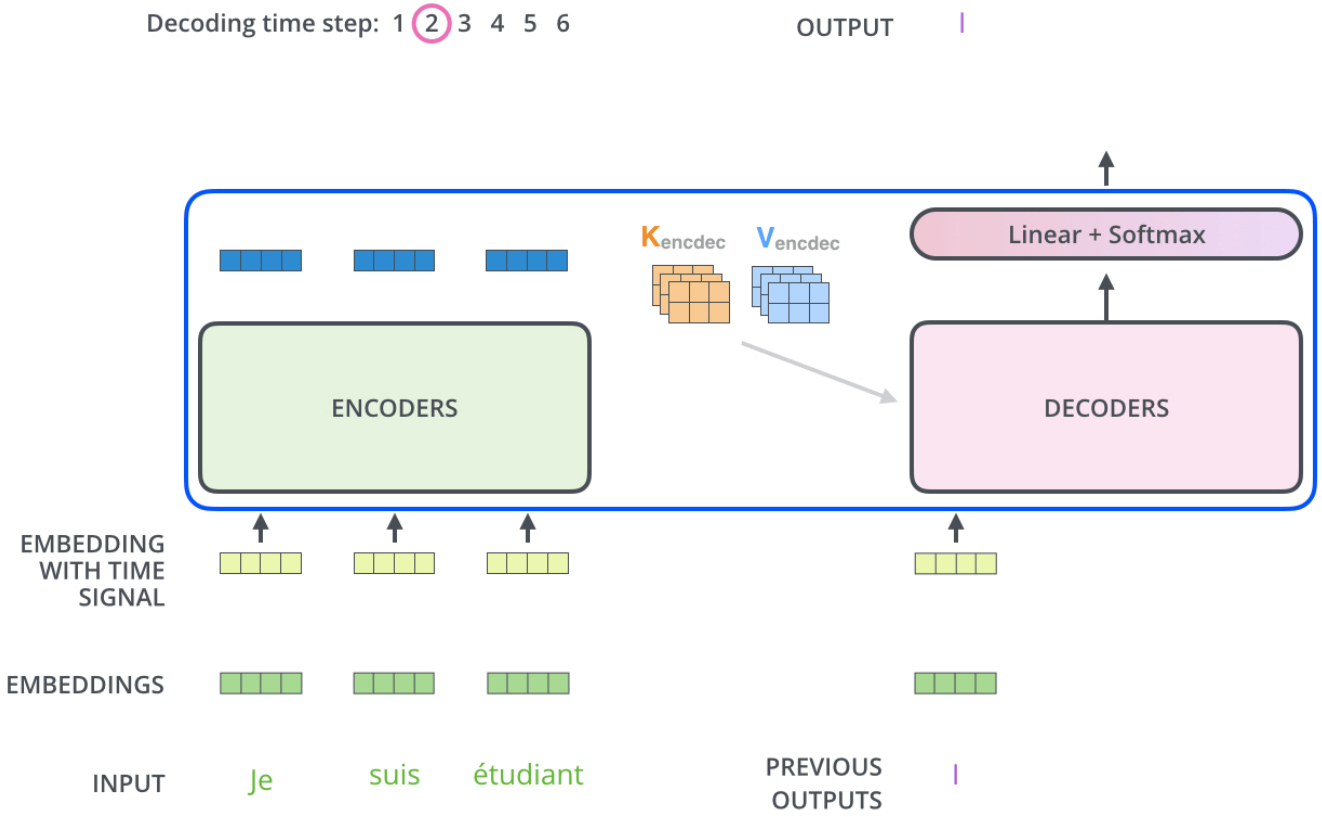


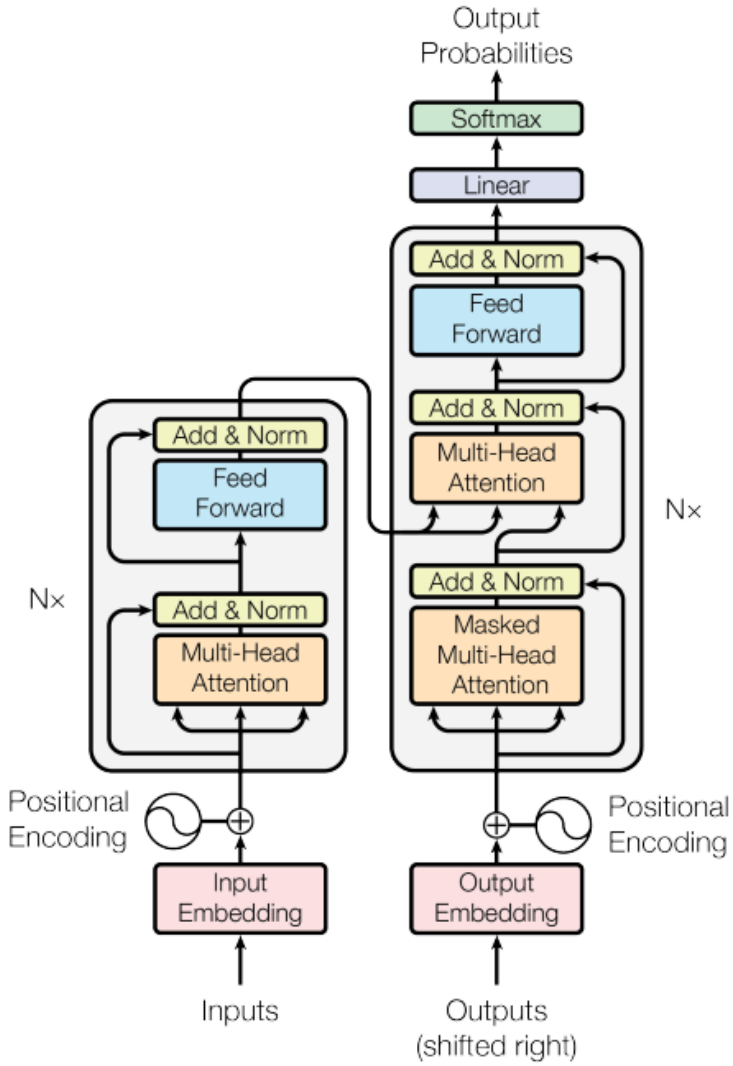


https://jalammar.github.io/illustrated-transformer/

Transformer

- Decoder





https://jalammar.github.io/illustrated-transformer/

BERT

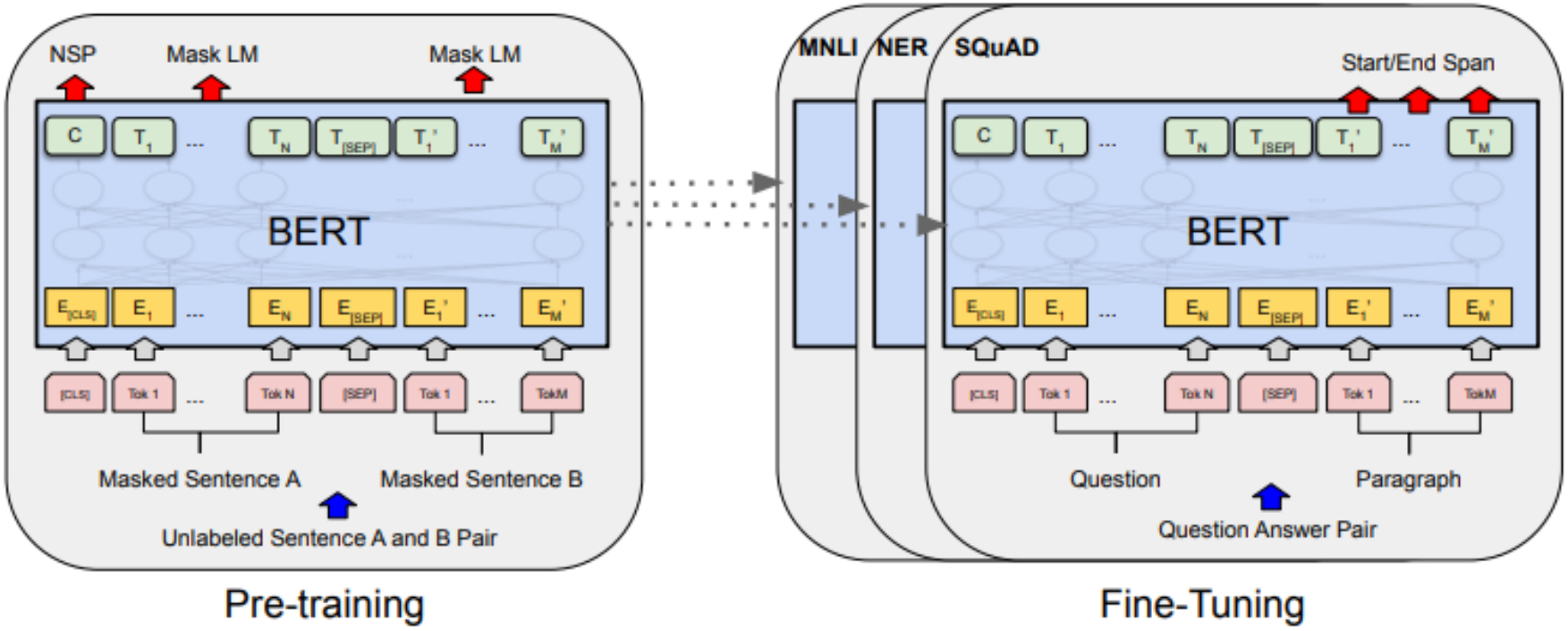
- **Quiz**
  - Minjoo _____ kimchi
    - ① run
    - ② read
    - ③ eat

BERT

- **Quiz**
  - Minjoo _____ kimchi
    - ① run
    - ② read
    - ③ **eat**

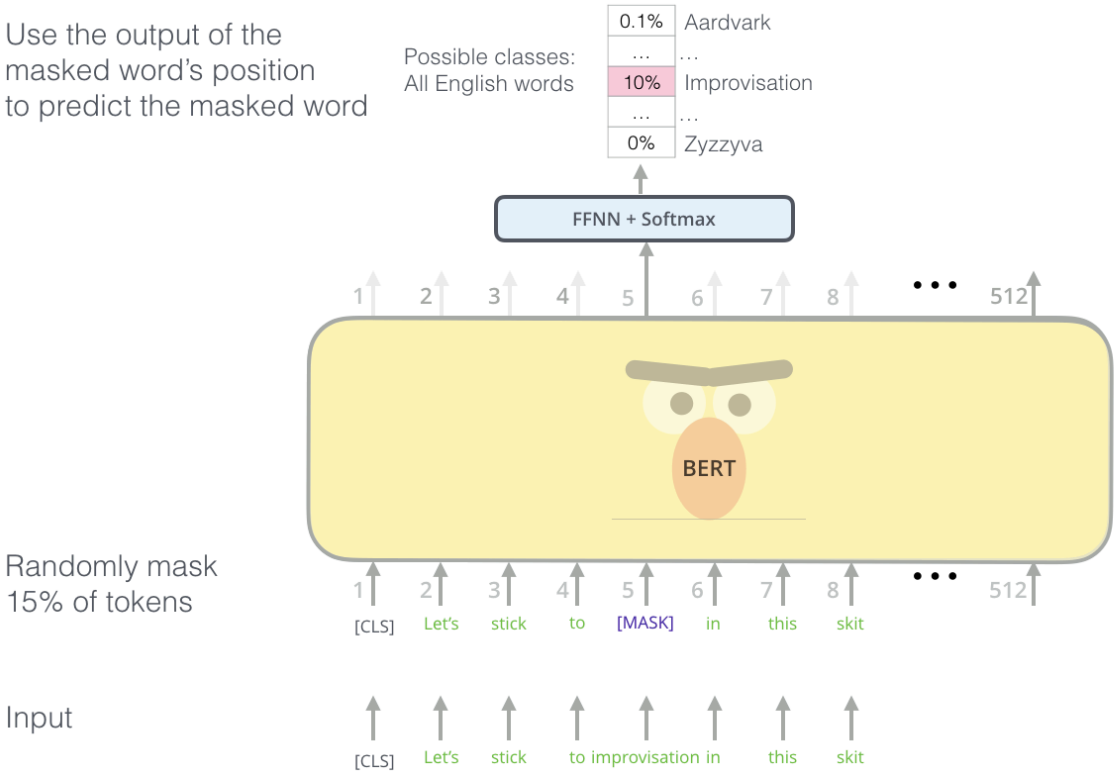
BERT

- **B**idirectional **E**ncoder **R**epresentations from **T**ransformers



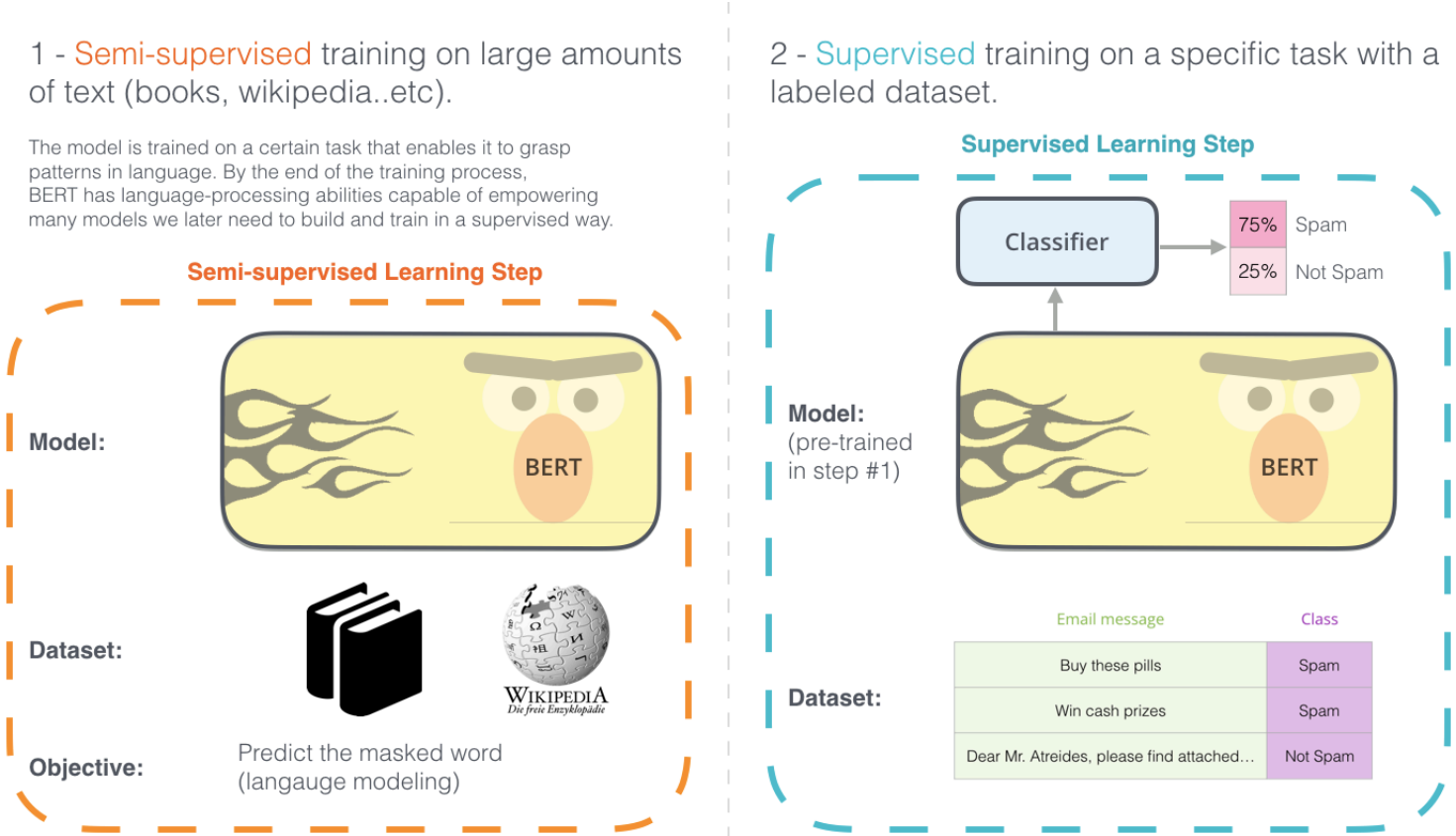
BERT

- Masked Language Model

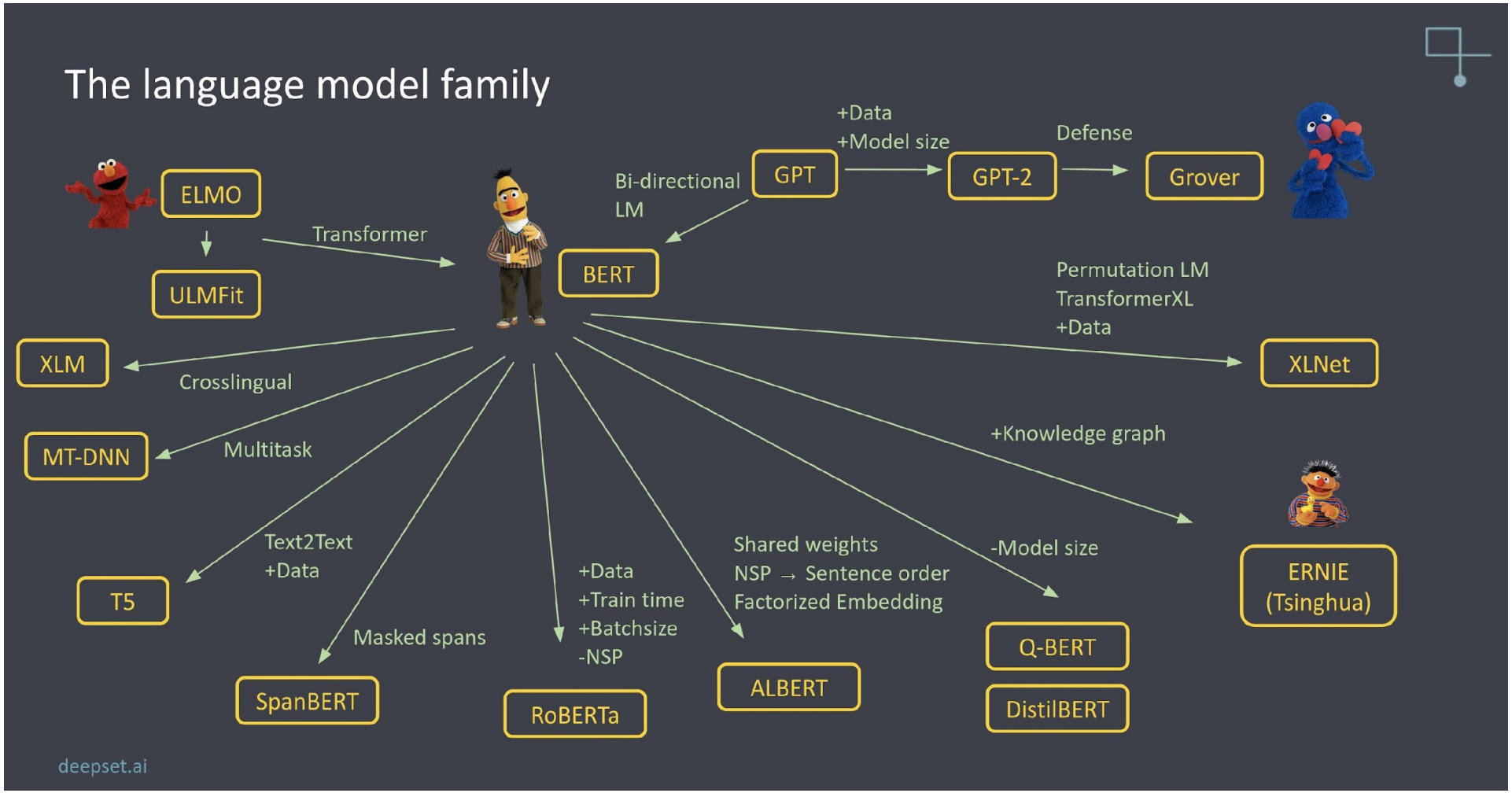


BERT

- Pre-trained model



Family of BERT



Machine Translation

Machine Translation?

- Process when a computer software translates text

  from one language to another without human involvement

History

RBMT

- Rule Based Machine Translation
- = PBMT (Phrase Based Machine Translation)

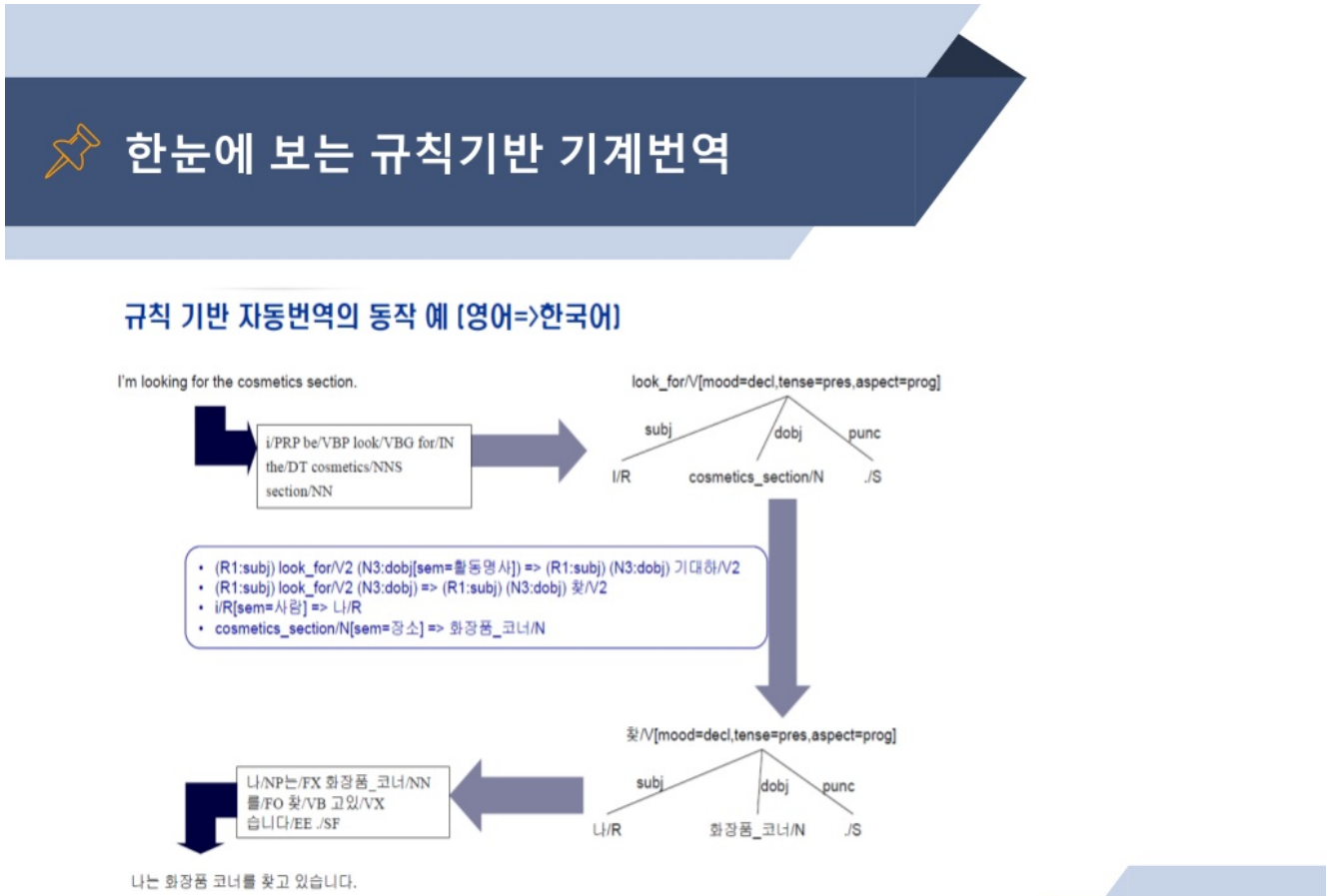
SMT

- Statistical Machine Translation

NMT

- Neural Machine Translation (by AI)
- Trained with sequence

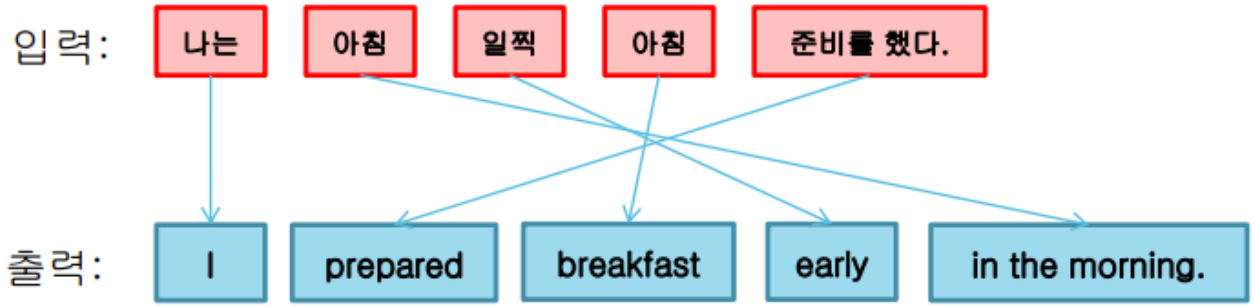
RBMT : Rule Based Machine Translation



https://www.slideshare.net/NaverEngineering/mt-173570380

SMT : Statistical Machine Translation

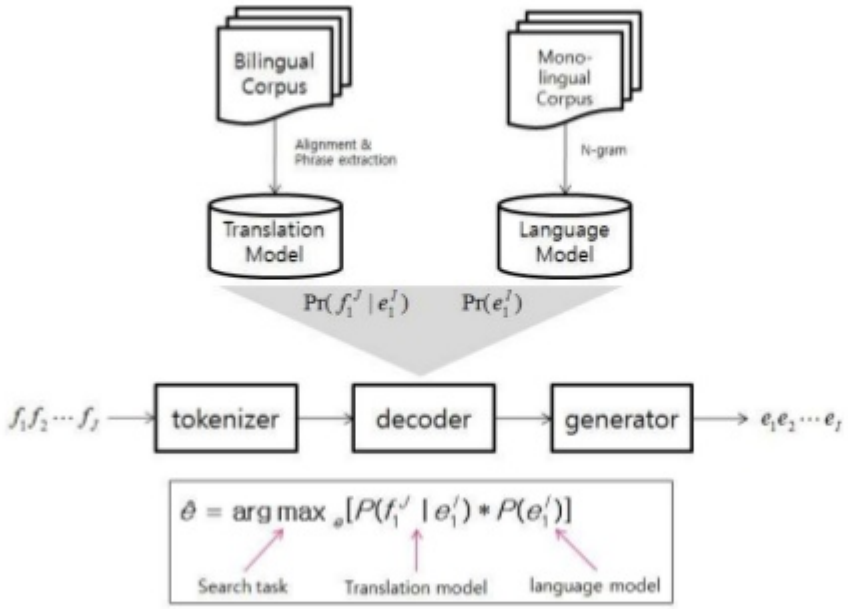
- 작동 방식



  - 입력 문장의 **일부분(Word 혹은 Phrase)**를 보고 번역 수행
    Translate using **word or phrase** of input sentence
  - **번역 사전**을 사용하여, 확률에 기반한 번역을 수행
    Statistical translation using a dictionary
- 어순이 같은 언어 (한국어/일본어, 영어/프랑스어)
  - 괜찮은 성능.

  Similar order language (Korean/Japanise, English/French) : **Not Bad**
- 어순이 다른 언어 (한국어/영어)
  - 성능 부족.

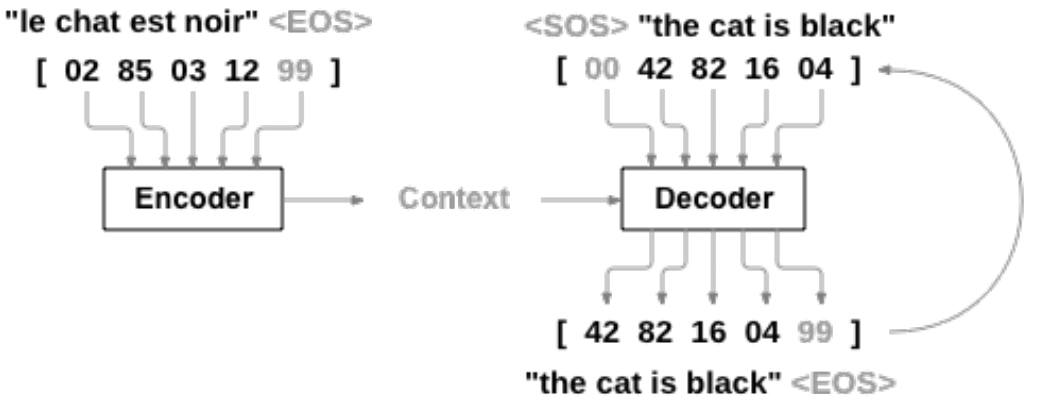
  Different order language (Korean/English) : **BAD**



https://www.slideshare.net/deview/papagon2mt

Neural Machine Translation (NMT)

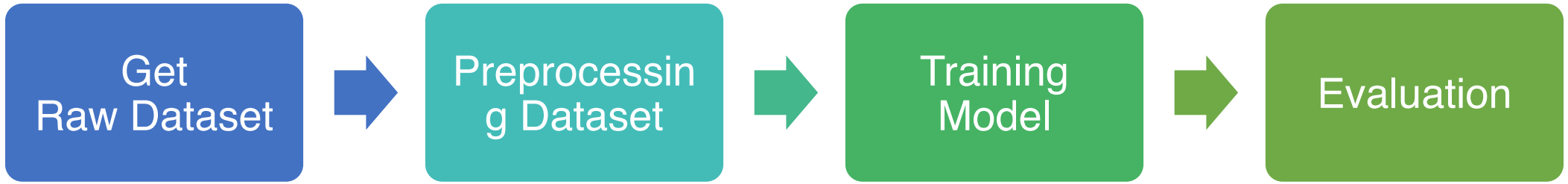
- Train translation model with language sentence pair
  - Dataset feature : [Source sentence] - [Target sentence]
    - Don’t need to build phrase analyzer / dictionary
    - Translate sentence by sentence -> Translate in context is possible



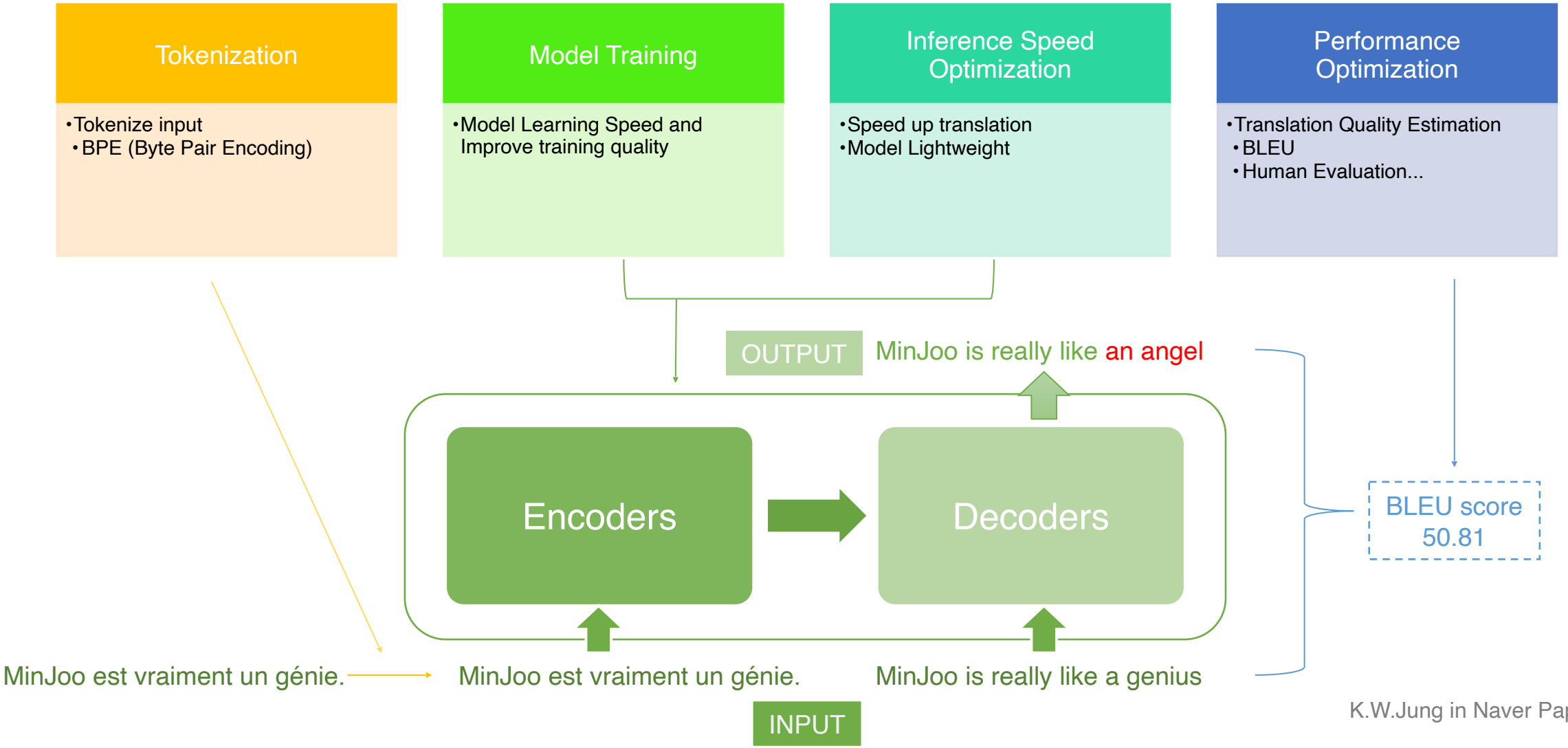
https://tutorials.pytorch.kr/_images/seq2seq.png

Process to Develop NMT

- NMT (Neural Machine Translation)



4 Elements of Machine Translation



K.W.Jung in Naver Papago

(1/4) Tokenization

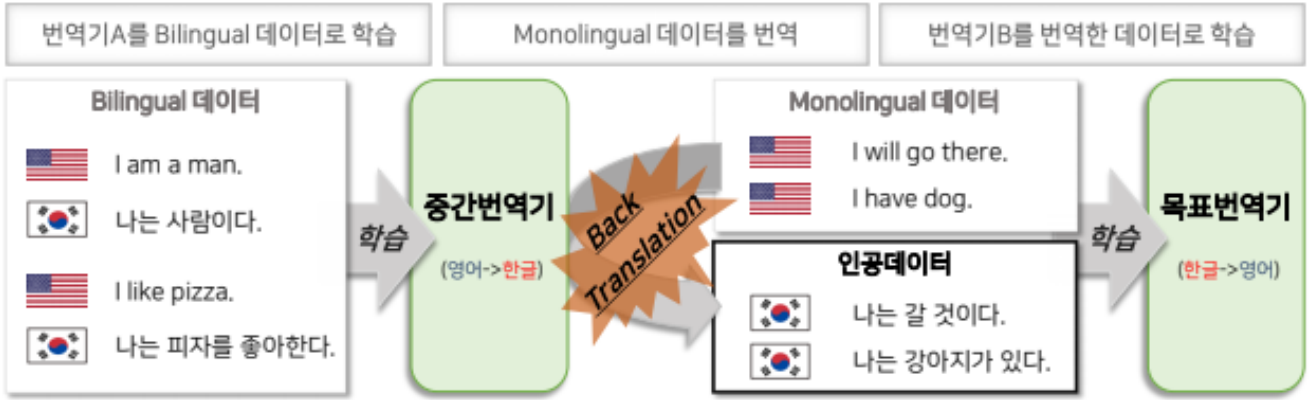
- Tokenize input sentence
  - 내용이 정말 거지같다. → 내용/NNG 이/JKS 정말/MAG 거지/NNG 같/VA 다/EF ./SF
  - Tokenizer example : (KOR) Mecab, Okt / (EN) Moses / (ZH) Jieba / ...

- BPE (Byte Pair Encoding)
  - Subword segmentation algorithm
  - OOV (Out Of Vocabulary) problem solved.
    - lowest -> low + est
  - https://github.com/rsennrich/subword-nmt

(2/4) Model Training

- Model
  - Fast model training strategy / High-quality strategy

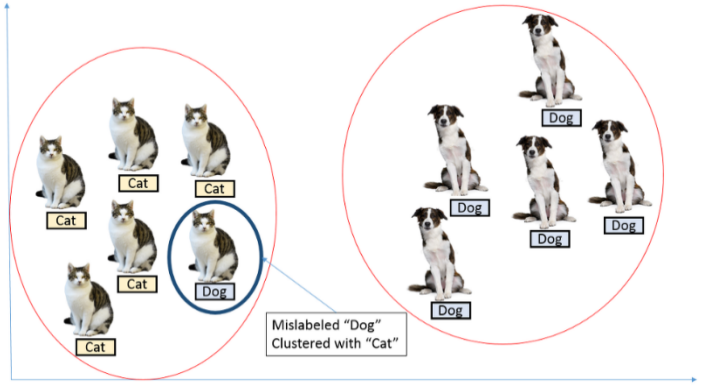
- Dataset
  - Data Augmentation
    - Back-Translation

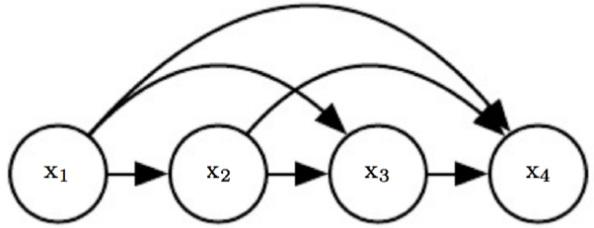


http://dsba.korea.ac.kr/seminar/?mod=document&uid=1328

(2/4) Model Training

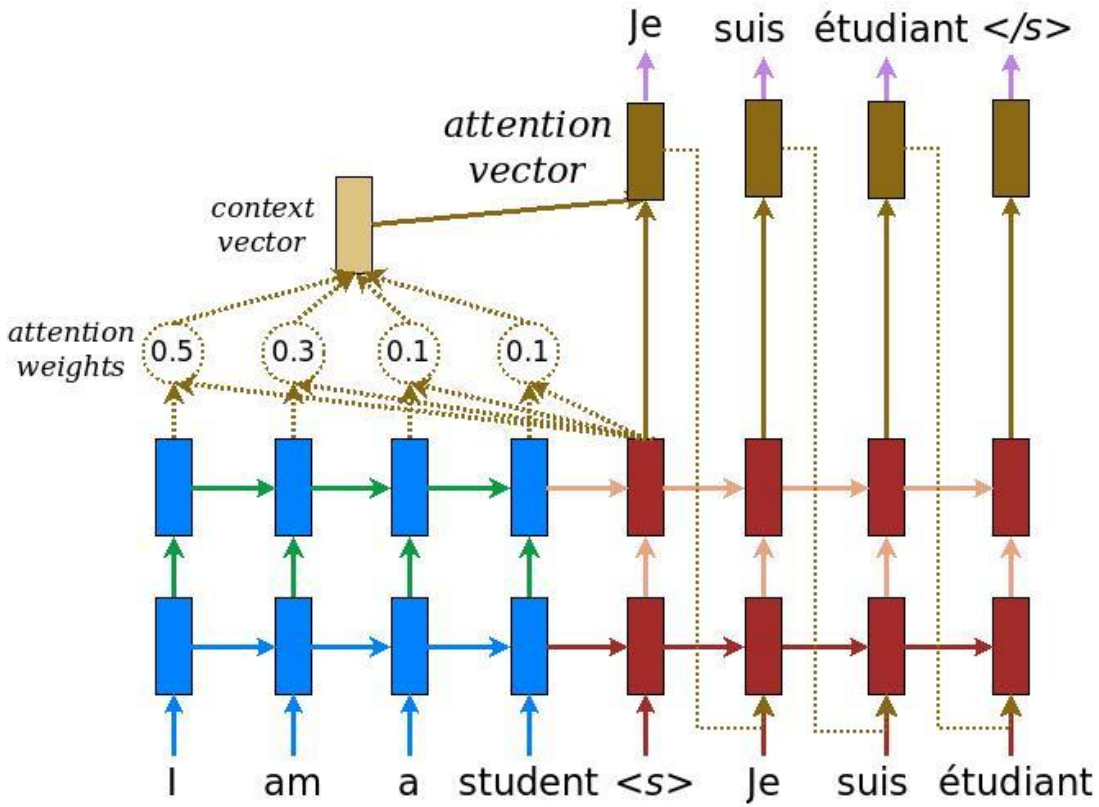
- Optimization
  - Label Smoothing
  - Curriculum learning
- Modeling
  - Non-Autoregressive
  - LSTM (Seq2Seq + Attention) → Transformer (Self-Attention)





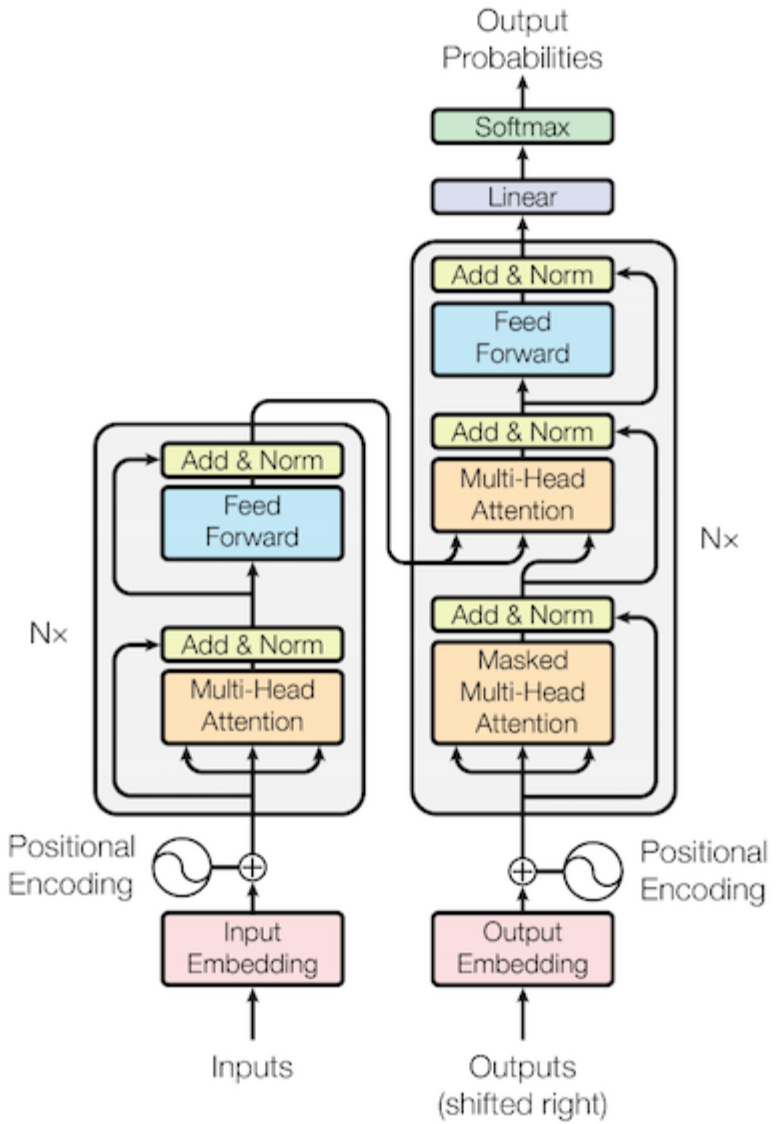
https://3months.tistory.com/465
https://ratsgo.github.io/generative%20model/2018/01/31/AR/

(2/4) Model Training



Seq2Seq+ Attention

https://www.tensorflow.org/text/tutorials/nmt_with_attention



Transformer (Self-Attention)

https://paul-hyun.github.io/transformer-03/

(3/4) Performance Optimization

- Measure performance of translation models
- Evaluate translation quality
    - Compare translation results with reference
- **BLEU**, ROUGE, RIBES...
    - Research about evaluation method better than BLEU
- More fast, reliable Human Evaluation

(3/4) Performance Optimization

- BLEU score (Bilingual Evaluation Understudy Score)
  - Evaluation method by N-gram

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}.$$

  - Moses multi-bleu perl
    - https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl

```python
# 1 word different
from nltk.translate.bleu_score import sentence_bleu
reference = [['the', 'quick', 'brown', 'fox', 'jumped', 'over', 'the', 'lazy', 'dog']]
candidate = ['the', 'fast', 'brown', 'fox', 'jumped', 'over', 'the', 'lazy', 'dog']
score = sentence_bleu(reference, candidate)
print("{:.2f}".format(score))
```

```
0.75
```

```python
# 2 word different
from nltk.translate.bleu_score import sentence_bleu
reference = [['the', 'quick', 'brown', 'fox', 'jumped', 'over', 'the', 'lazy', 'dog']]
candidate = ['the', 'fast', 'brown', 'fox', 'jumped', 'over', 'the', 'sleepy', 'dog']
score = sentence_bleu(reference, candidate)
print("{:.2f}".format(score))
```

```
0.49
```

https://machinelearningmastery.com/calculate-bleu-score-for-text-python/

(4/4) Inference Speed Optimization

- Improve translation speed
  - GPU inference, CPU inference

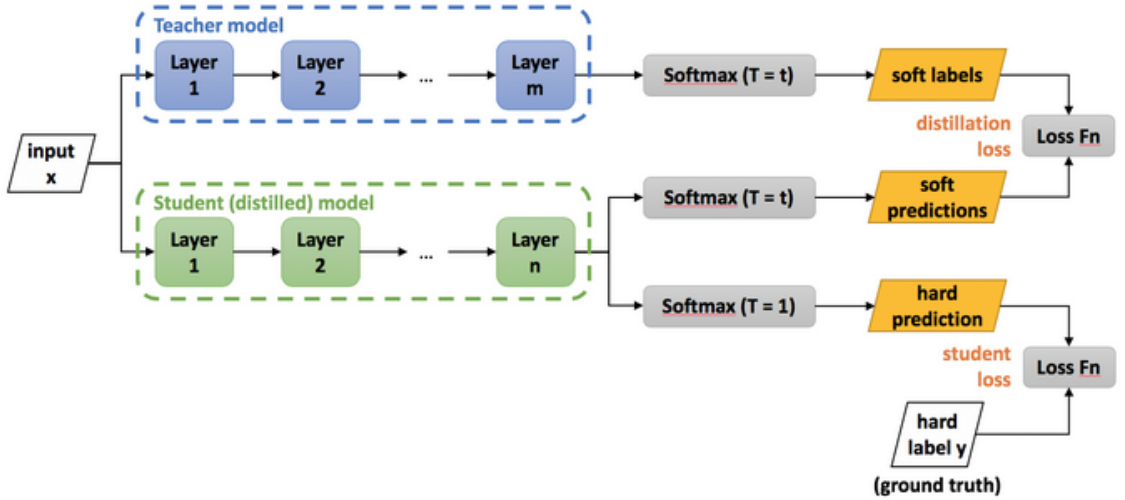
- Lighten Model, On-Device NMT
  - Decrease Encoder or Decoder Layer of NMT
    - Decrease model size, increase speed
  - Decrease model capacity (Algorithm) or change model structure
  - Decrease or share parameter
  - ※ Often a slight decrease in translation performance (trade-off)

(4/4) Inference Speed Optimization

- (Additional) Knowledge Distillation
  - Transfer pre-trained knowledge (softer softmax)

    from teacher (big model)

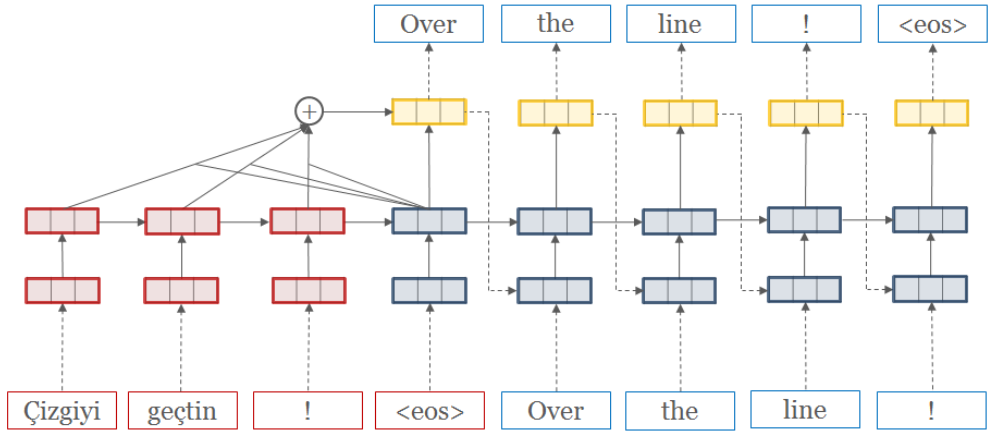
    to student (small capacity model)



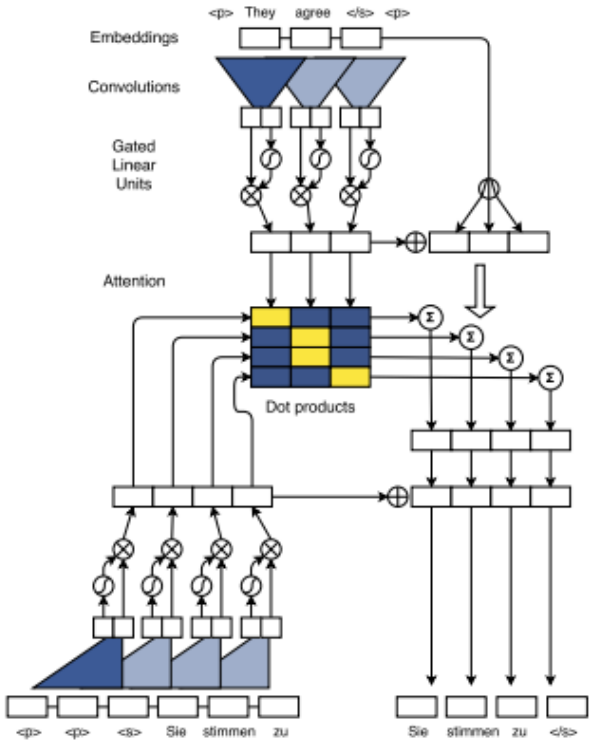
- https://light-tree.tistory.com/196
- https://intellabs.github.io/distiller/knowledge_distillation.html
- http://dmqm.korea.ac.kr/activity/seminar/304
- https://blog.lunit.io/2018/03/22/distilling-the-knowledge-in-a-neural-network-nips-2014-workshop/

Open Source

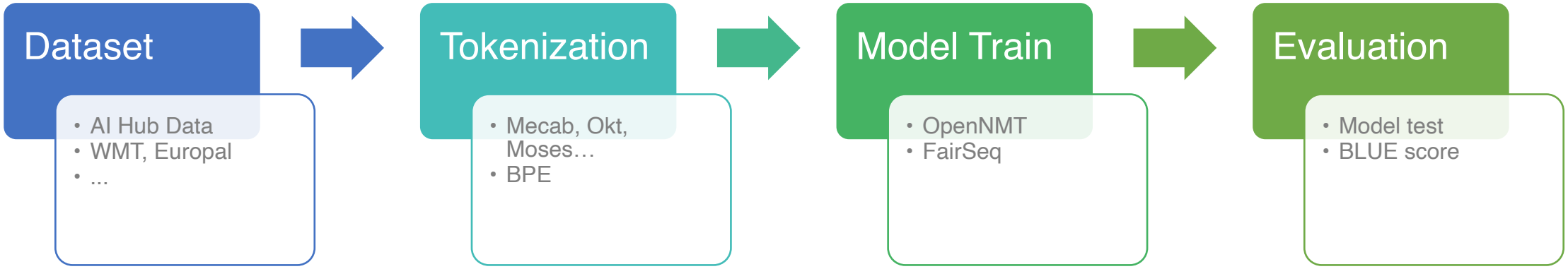
- OpenNMT
  - by Havard NLP group & Systran
  - tensorflow, pytorch



- FairSeq (**F**acebook **AI** **R**esearch **Seq**uence-to-Sequence Toolkit)
  - by Facebook AI Research
  - pytorch

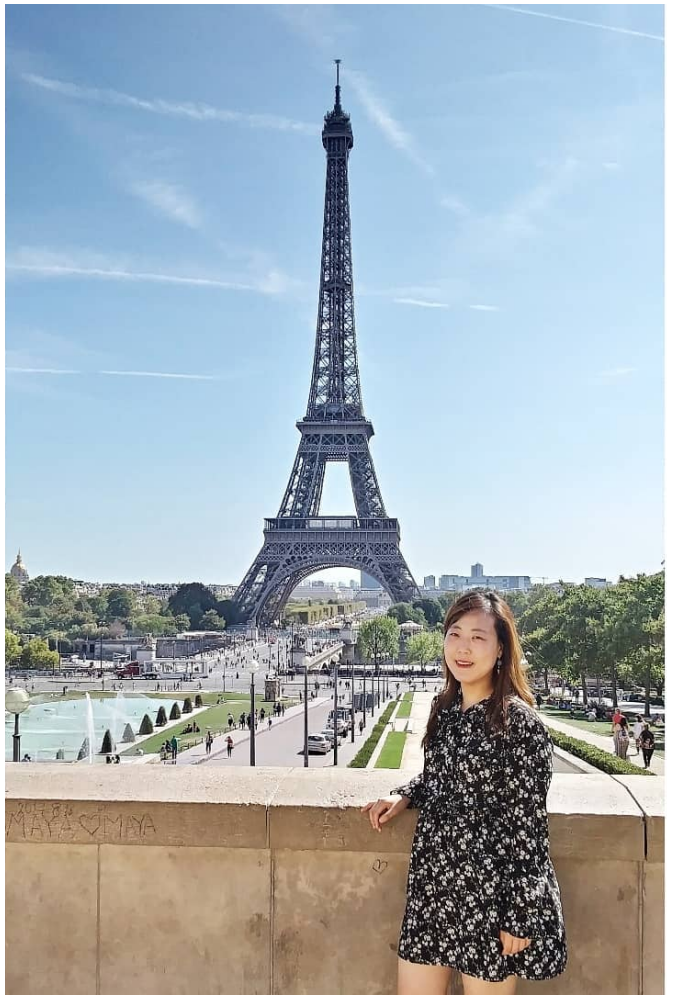


Example



Original (KO)	Reference (EN)	OpenNMT Translation Result
우리는 일주일에 5 일을 학교에 갑니다.	We go to school 5 days a week.	We go to school for 5 days a week.
다음 주말에는 함께 술을 마시기로 해요.	Let's drink together next weekend.	I will drink together next weekend.
난 당신이 행복해지기를 너무나 원해요.	I so want you to be happy.	I really want you to be happy.

Q & A



https://www.facebook.com/minjoo.choi.562/

https://github.com/Judy-Choi