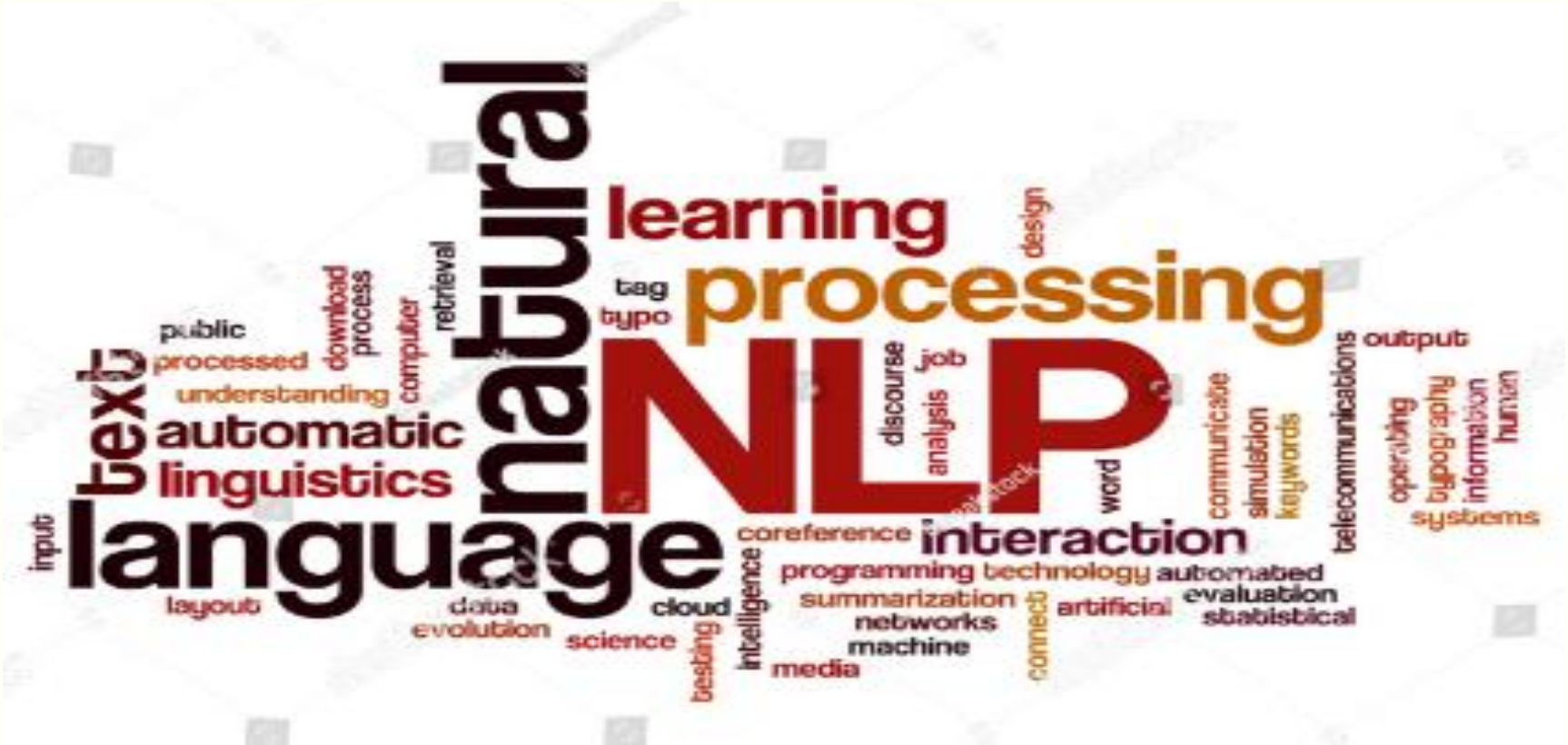
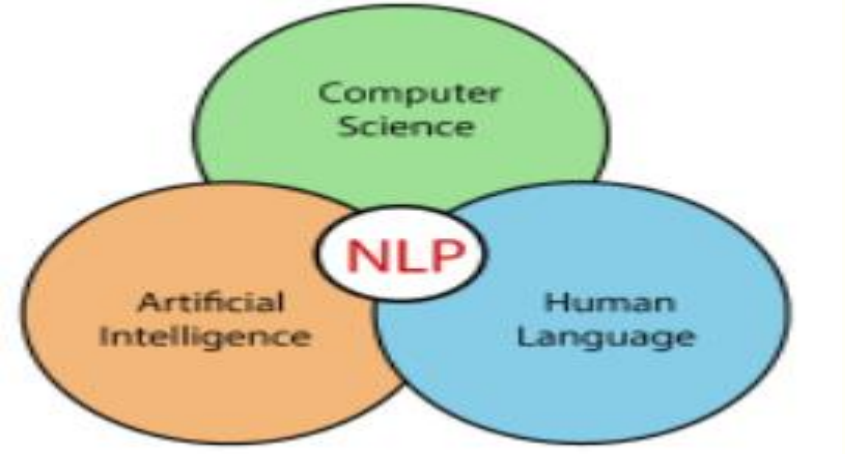
NATURAL LANGUAGE PROCESSING



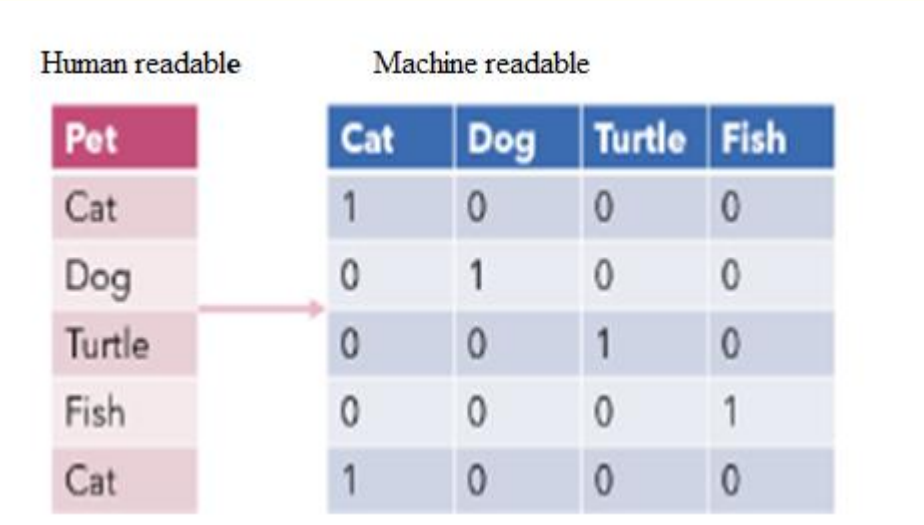
INTRODUCTION TO NLP

Natural Language Processing (NLP) is itself a broad field that lies under Artificial Intelligence. NLP depends upon linguistics and is responsible for making computers understand the text and spoken words the same way humans do.



How Does a Computer Understand Text

Word representation is a technique to represent a word with a vector and each word has its unique vector representation. One-hot encoding is one such technique used to convert categorical data into numerical data.



Advantages

1. . NLP helps users to ask questions about any subject and get a direct response within seconds.
2. NLP offers exact answers to the question means it does not offer unnecessary and unwanted information.
3. NLP helps computers to communicate with humans in their languages.
4. It is very time efficient.
5. Most of the companies use NLP to improve the efficiency of documentation processes, accuracy of documentation, and identify the information from large databases.

Disadvantages

1. NLP may not show context.
2. NLP is unpredictable
3. NLP is unable to adapt to the new domain, and it has a limited function that's why NLP is built for a single and specific task only.

Components of NLP

Natural Language Understanding (NLU)

Helps the machine to understand and analyse human language by extracting the metadata from content such as concepts, entities, keywords, emotion, relations, and semantic roles.

Natural Language Generation (NLG)

Acts as a translator that converts the computerized data into natural language representation. It mainly involves Text planning, Sentence planning, and Text Realization.

APPLICATION AREAS

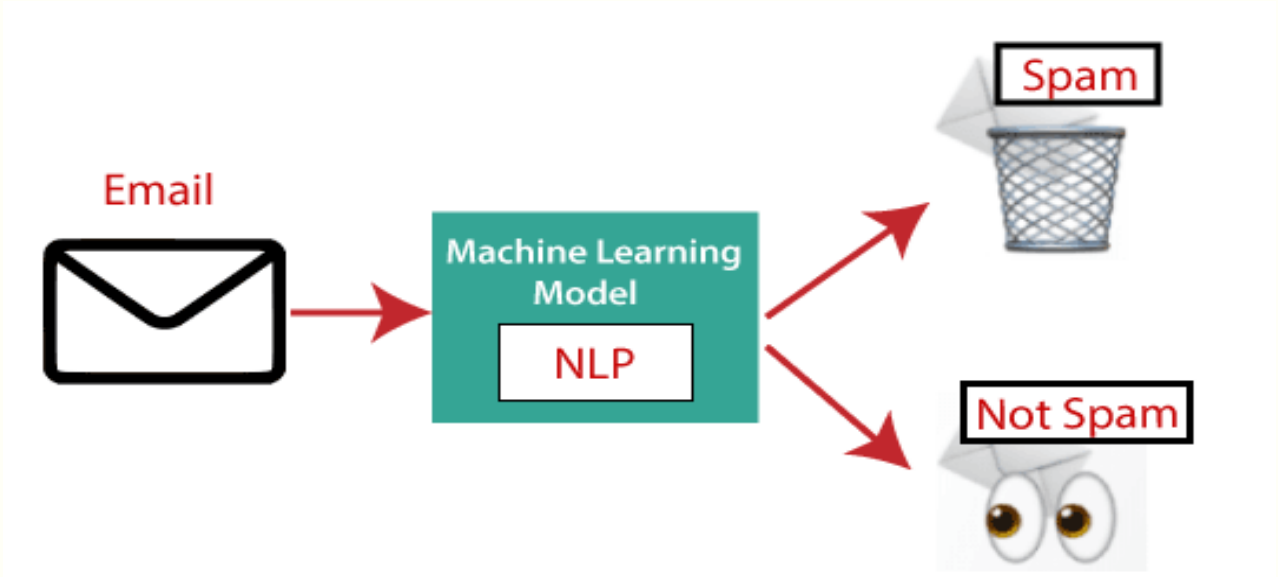
Question Answering

Question Answering focuses on building systems that automatically answer the questions asked by humans in a natural language.

APPLICATION AREAS

Spam Detection

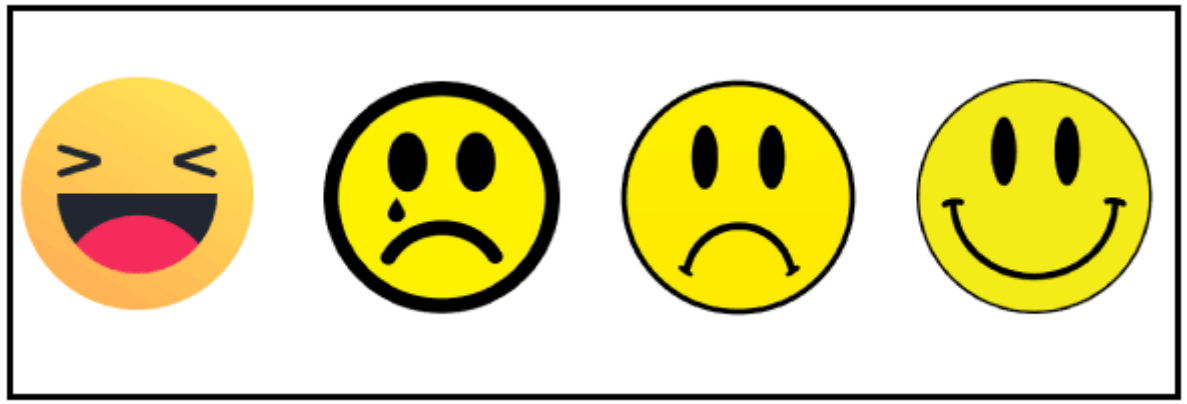
Spam detection is used to detect unwanted e-mails getting to a user's inbox.



APPLICATION AREAS

Sentiment Analysis

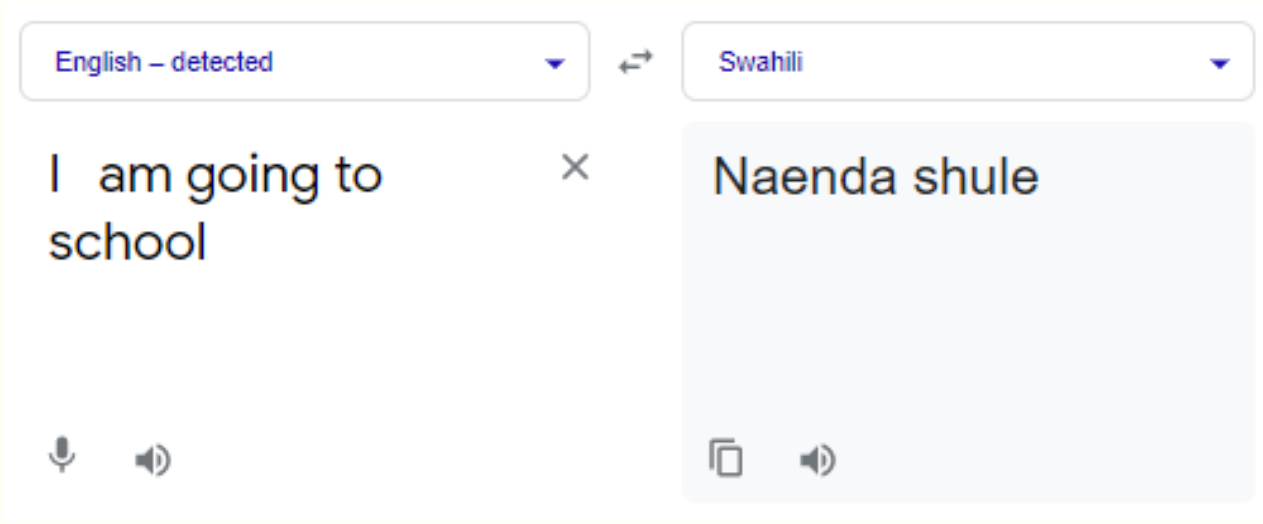
Sentiment Analysis is also known as **opinion mining**. It is used on the web to analyse the attitude, behaviour, and emotional state of the sender.



APPLICATION AREAS

Machine Translation

Machine translation is used to translate text or speech from one natural language to another natural language.



APPLICATION AREAS

Spelling correction

Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling correction.

APPLICATION AREAS

Speech Recognition

Speech recognition is used for converting spoken words into text. It is used in applications, such as mobile, home automation, video recovery, dictating to Microsoft Word, voice biometrics, voice user interface, and so on.

APPLICATION AREAS

Chatbot

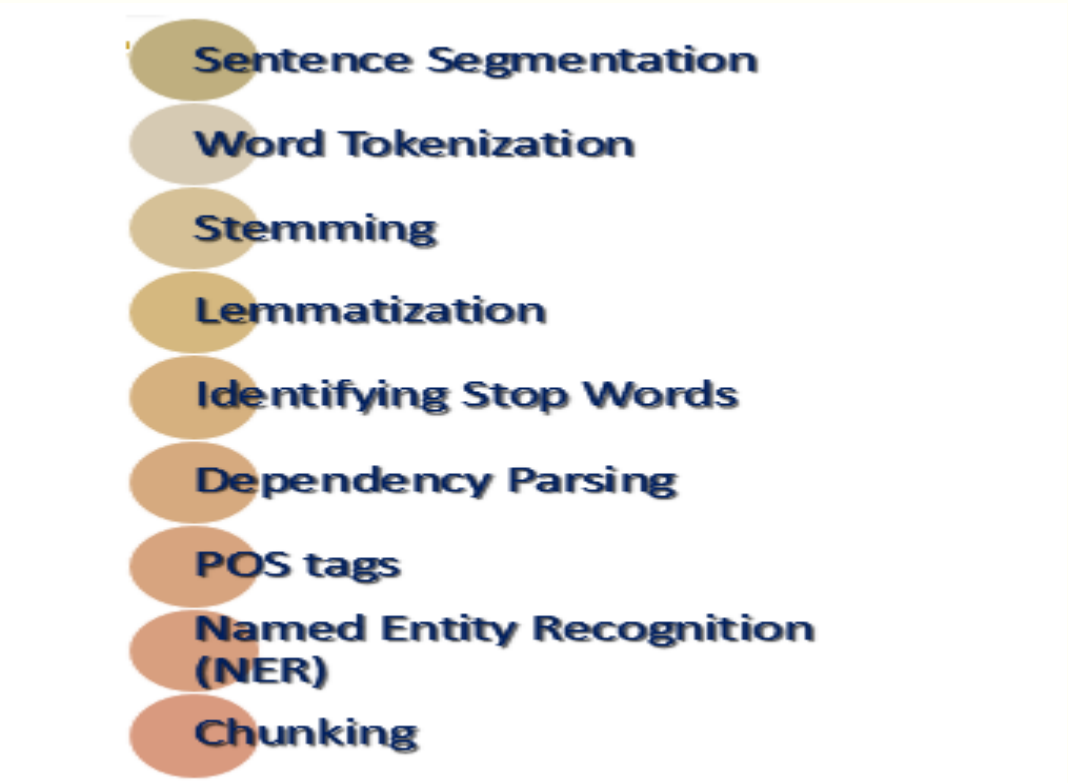
Implementing the Chatbot is one of the important applications of NLP. It is used by many companies to provide the customer's chat services.

APPLICATION AREAS

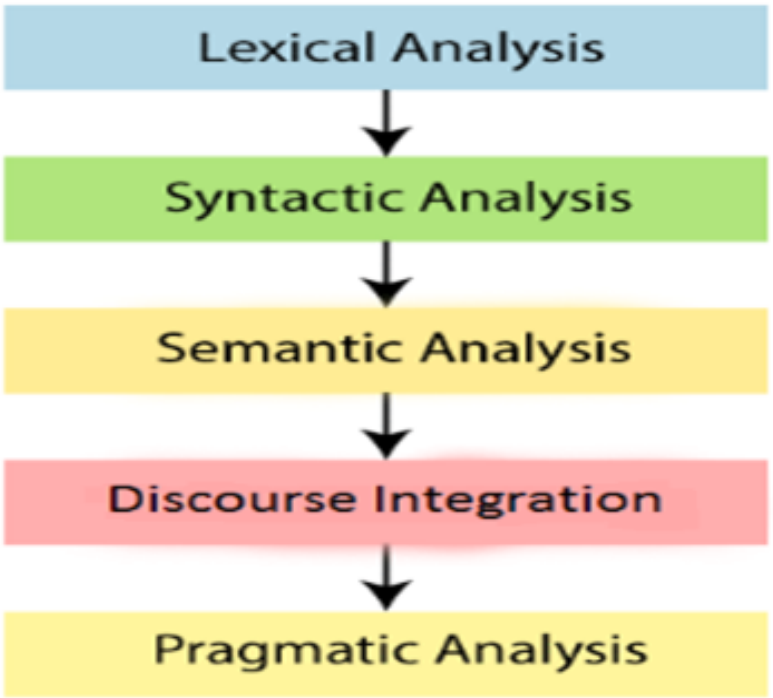
Information extraction

Information extraction is one of the most important applications of NLP. It is used for extracting structured information from unstructured or semi-structured machine-readable documents.

Building NLP pipeline steps



NLP phases



Identifying the Structure of words

Analysing sentence for grammar

Draws exact meaning of words

Words arrangement before and after matters

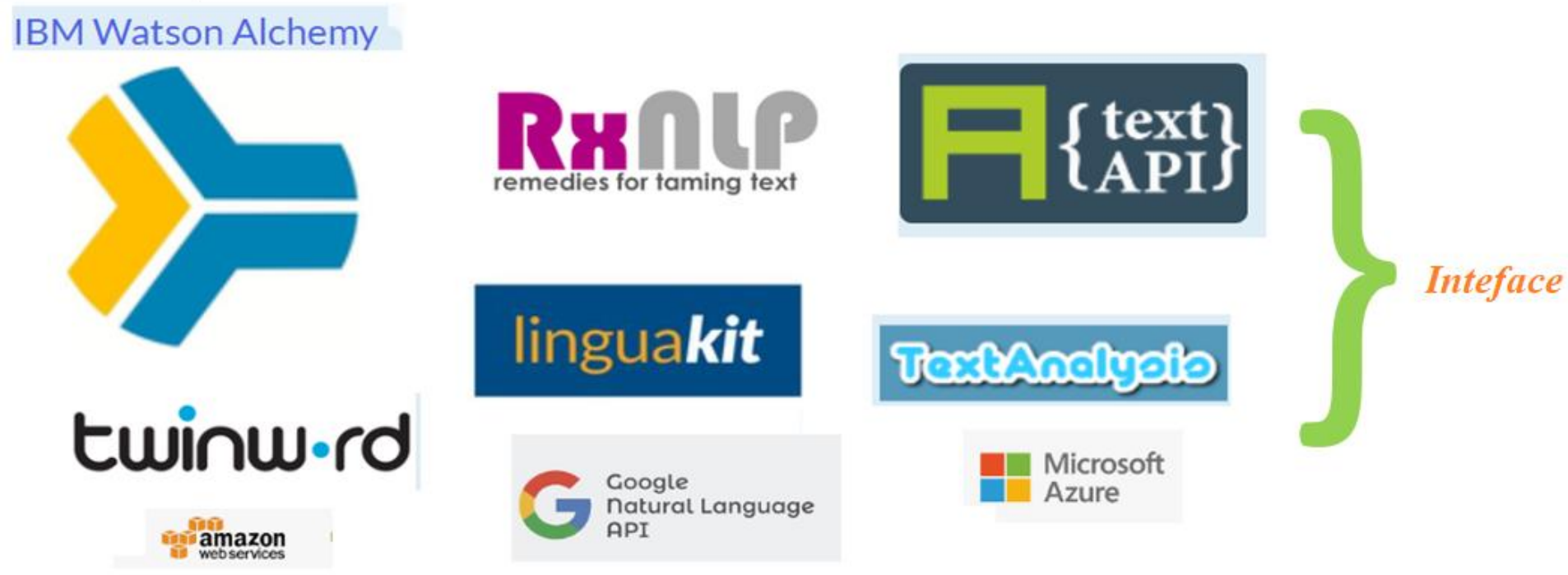
Derive aspects of real world knowledge

NLP Ambiguity

- **Lexical Ambiguity** *''I went to the bank''*
- **Syntactic Ambiguity** *''I saw the girl with the binocular''.*
- **Referential Ambiguity** *''The boy told his father about the theft. He was very upset''*

.

NLP API


IBM Watson Alchemy
RxNLP
remedies for taming text
{text API}
linguakit
TextAnalysis
twinword
Google Natural Language API
Microsoft Azure
amazon web services
Inteface

NLP Libraries- coding

The Transformer Model

Transformer model for NLP

- A transformer is a deep learning model that adopts the mechanism of self **attention**, **transduction**, **differentially weighting** the significance of each part of the input data
- Transformers were developed to solve any task that transforms an input sequence to an output sequence. Such as translation , classification etc.
- It allows parallel processing , therefore faster than most models such as RNNs, LSTMs .

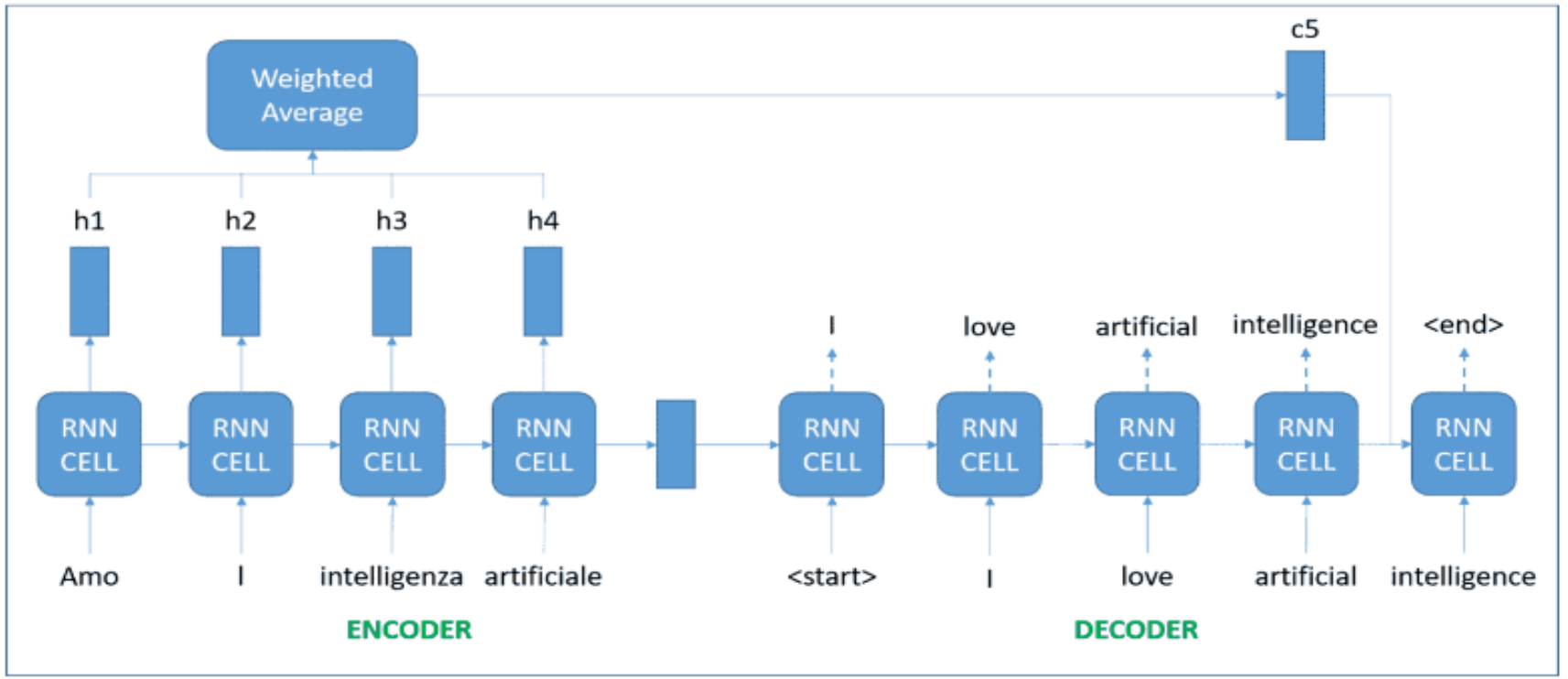
Translation…

Transformer architecture

Transformer model uses **Encoder- decoder** architecture

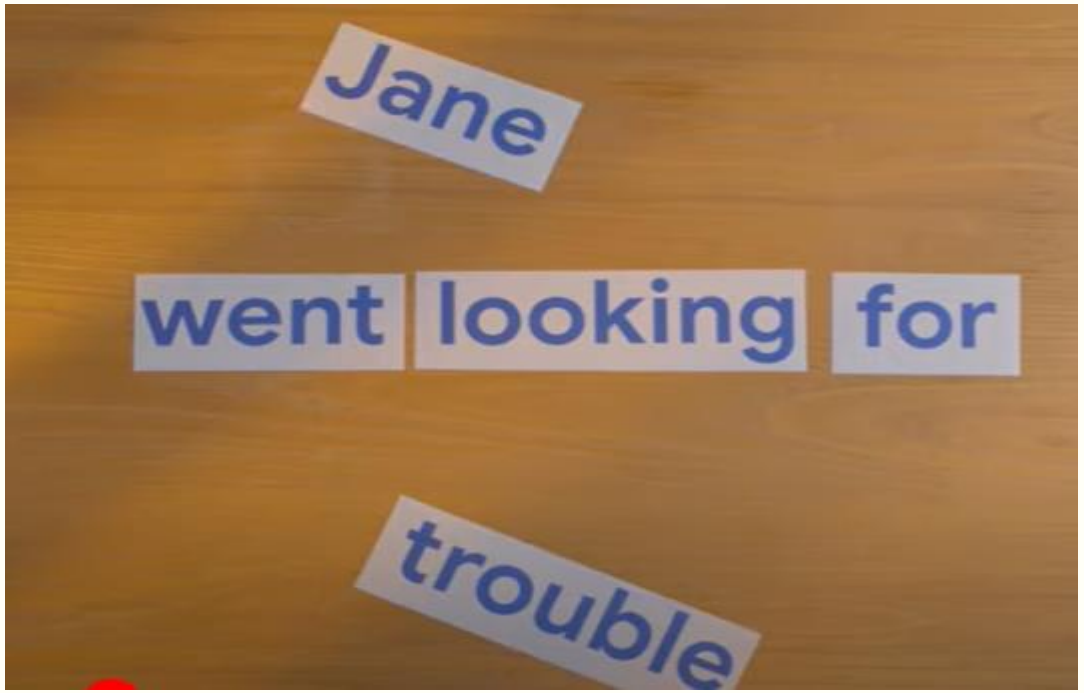
- The encoder consists of encoding layers that process the input iteratively- one layer after another
- Each encoder consists of two main layers: **Self-attention** and a feed **Forward Neural Network**.
- Decoder consist of decoding layers that uses features to produce output based on encoding scheme.

Translation..

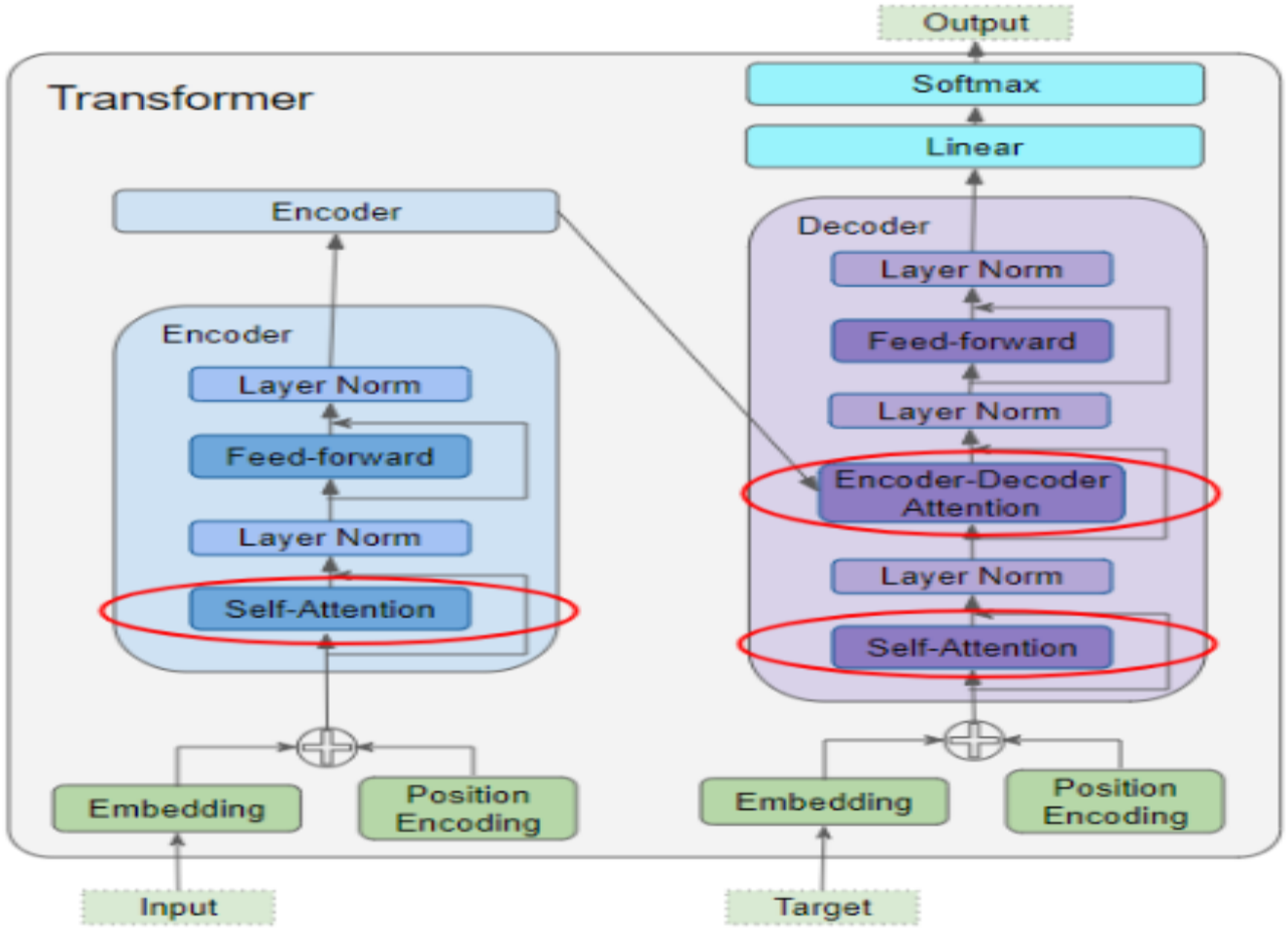


Tansformers for encoder decoder

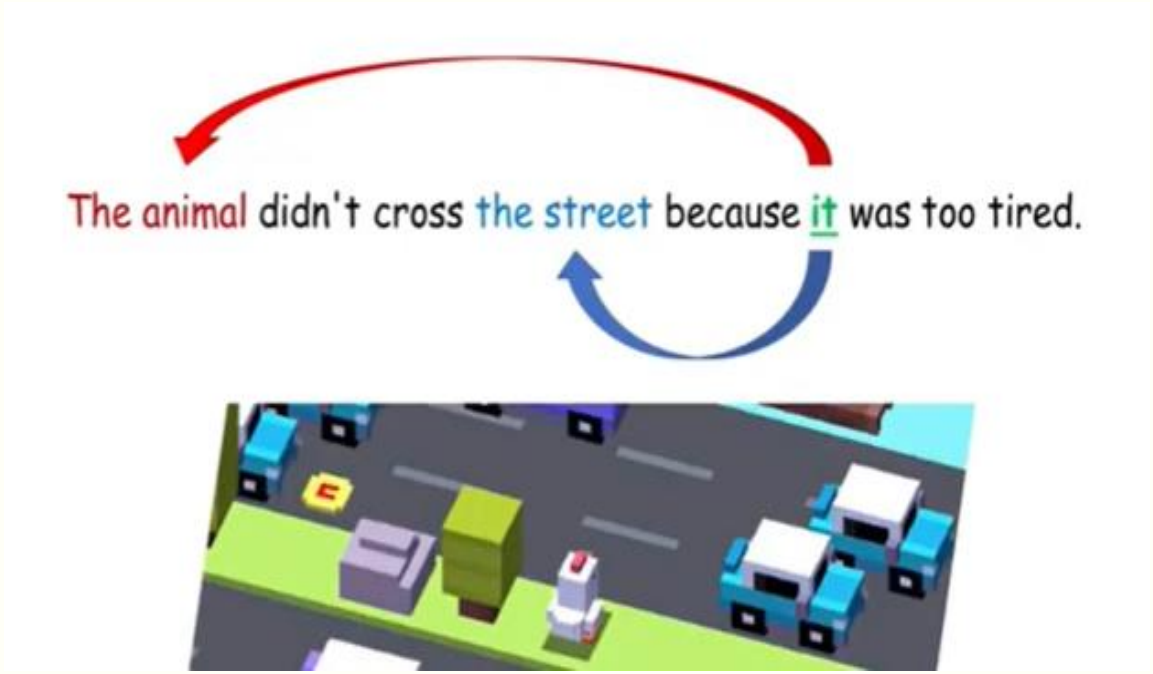
Order matters

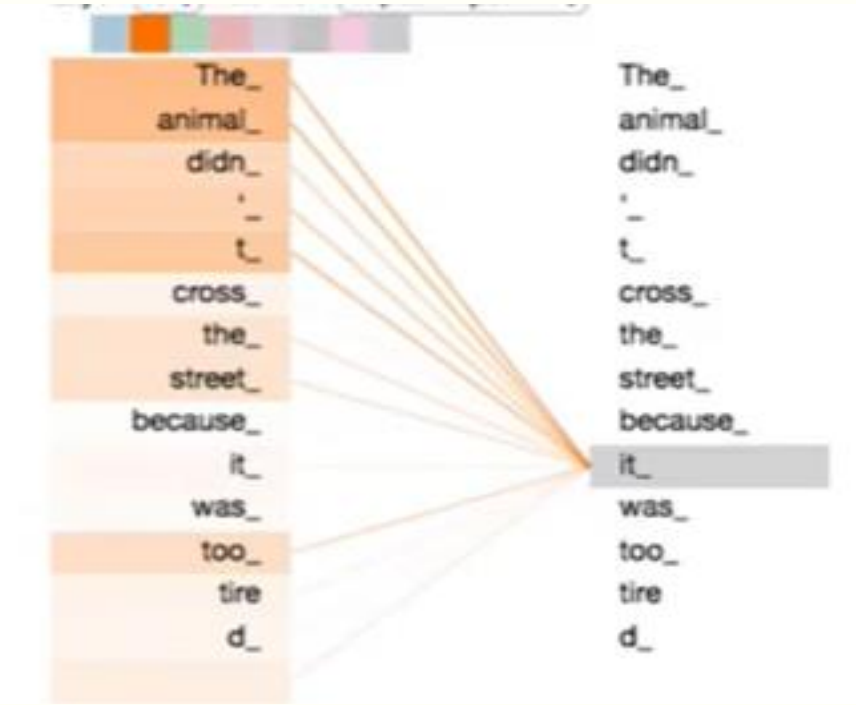


Transformer Architecture- The BEAST



Attention is all you need

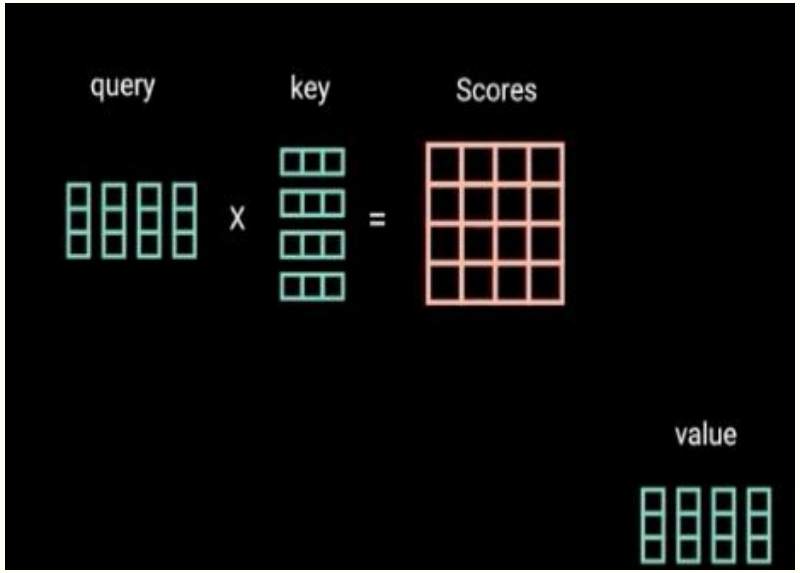




Query, Key, and Value Vectors (Q , K and V)

To achieve self-attention, we feed the input into 3 distinct fully connected layers to create the query, key, and value vectors.

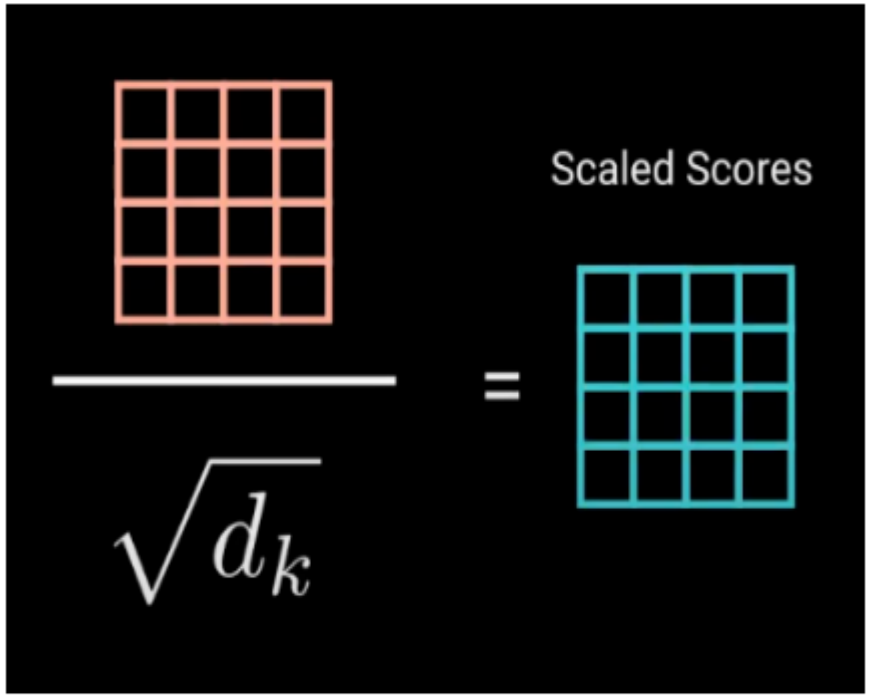
For example, when you type a query to search for some video on YouTube, the search engine will map your **query** against a set of **keys** (video title, description etc.) associated with candidate videos in the database, then present you the best matched videos (**values**).



Attention Vector

	Hi	how	are	you
Hi	98	27	10	12
how	27	89	31	67
are	10	31	91	54
you	12	67	54	92

Attention scores from the dot product.

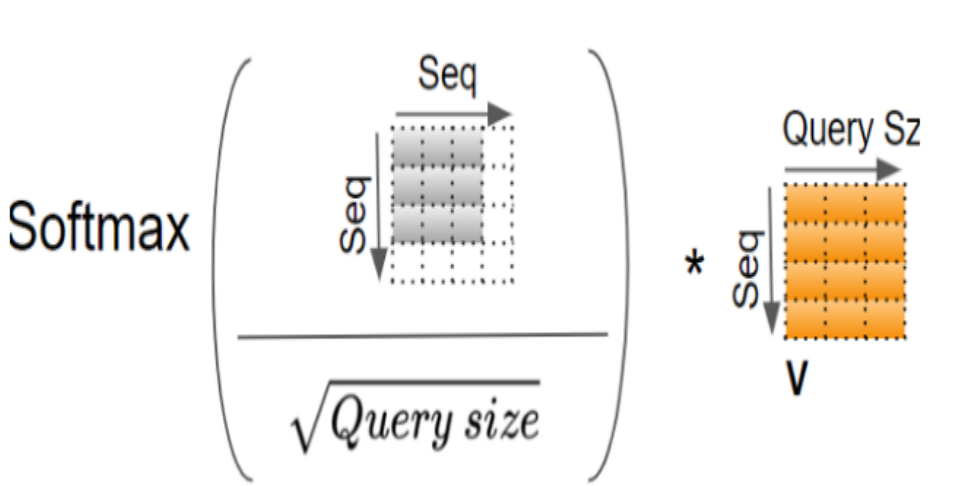


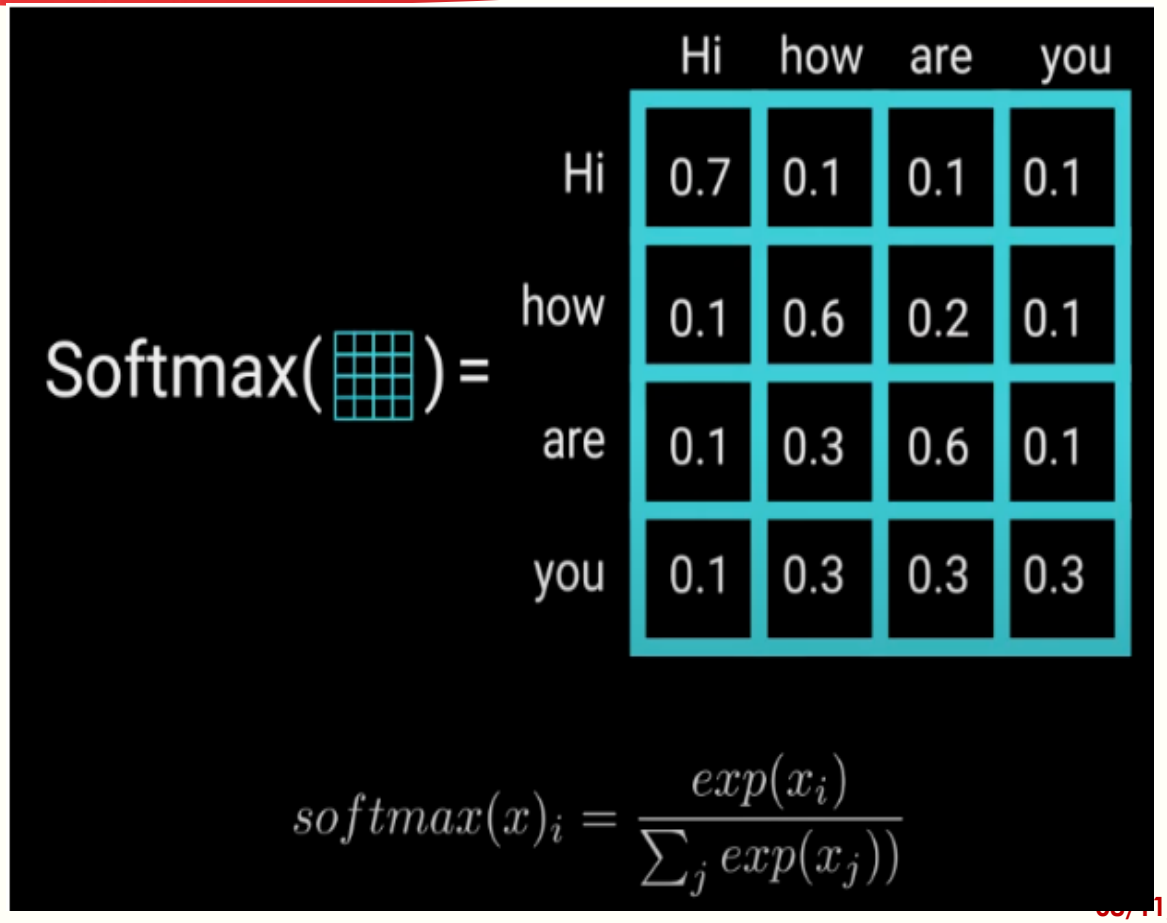
Scaling down the Attention scores

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Softmax for the Scaled Scores

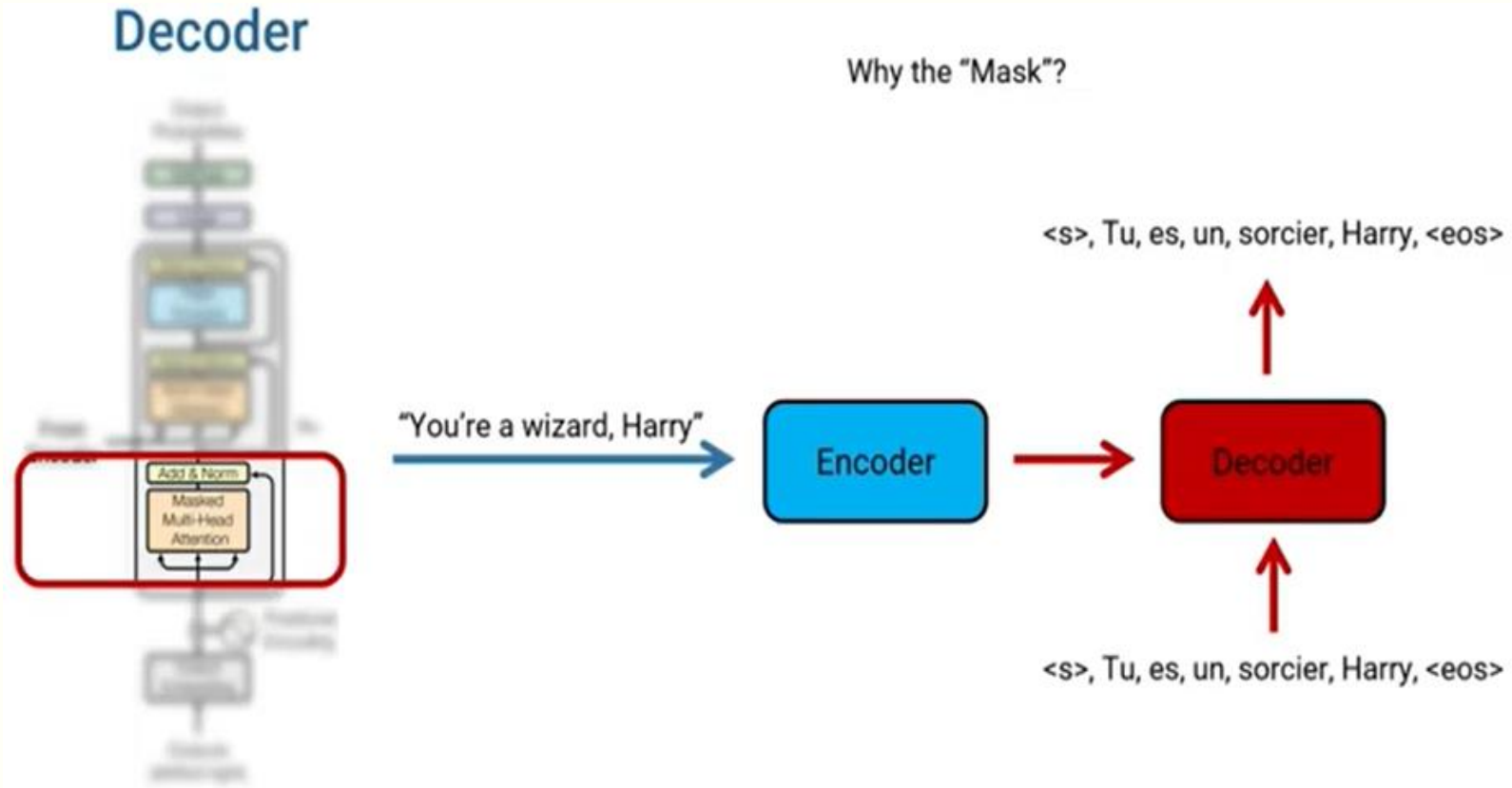
Softmax gives a probability values between 0 and 1. By doing a Softmax the higher scores get heighten, and lower scores are depressed. This allows the model to be more confident about which words to attend too.



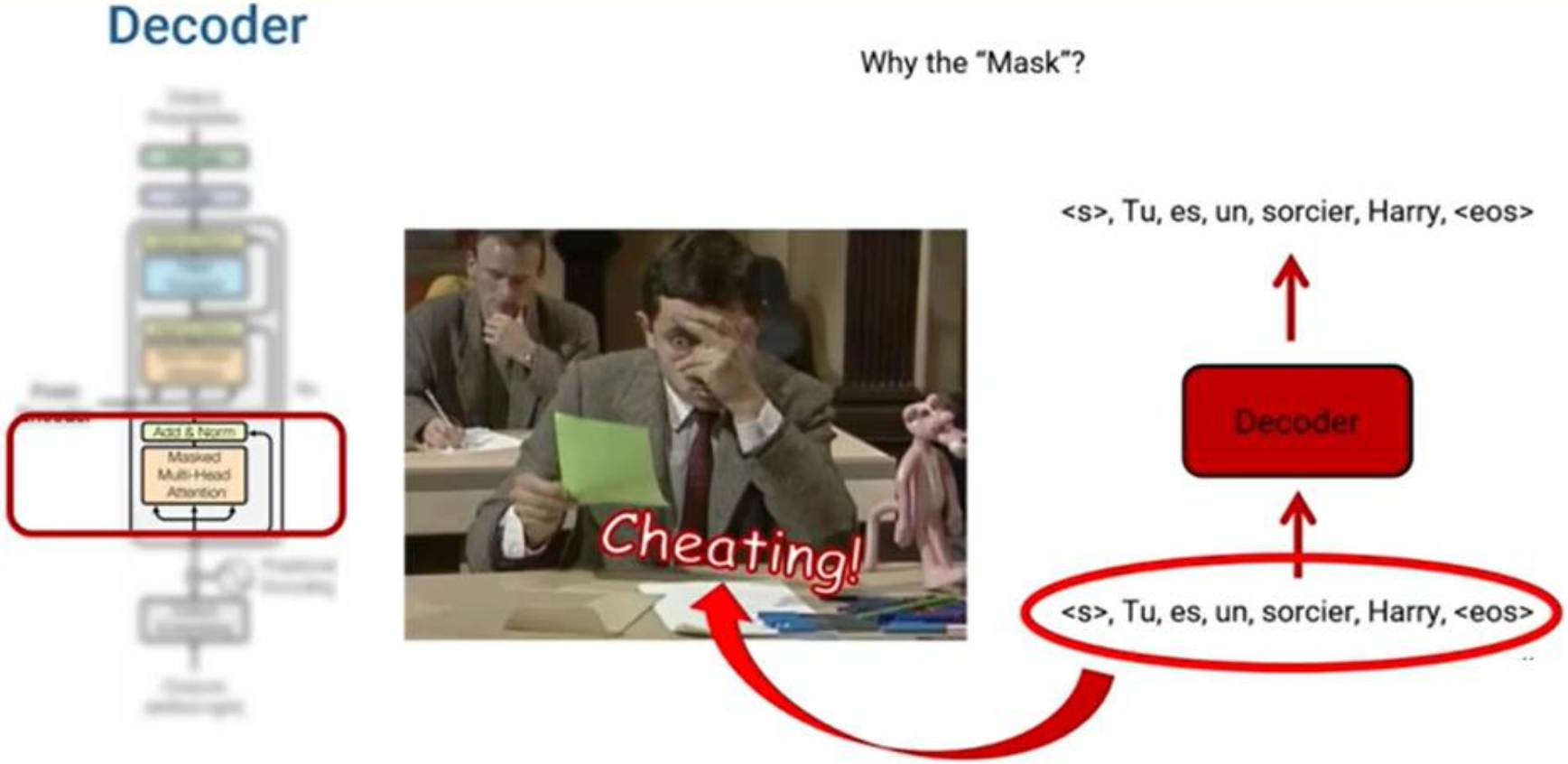


Masking

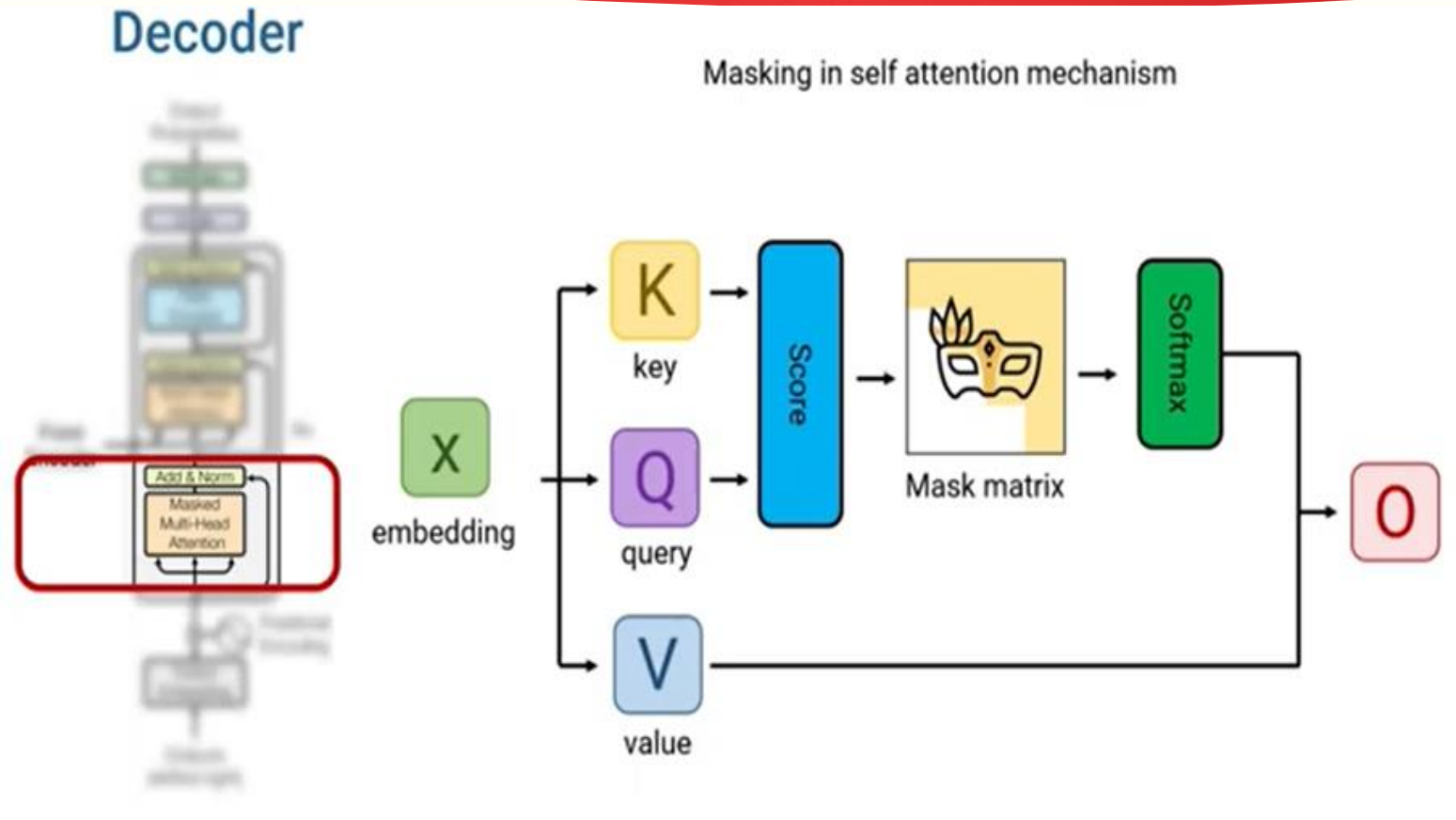
Masking


Decoder
Why the "Mask"?
Add & Norm
Masked Multi-Head Attention
"You're a wizard, Harry"
Encoder
Decoder
<s>, Tu, es, un, sorcier, Harry, <eos>
<s>, Tu, es, un, sorcier, Harry, <eos>

Masking


Decoder
Why the "Mask"?
Add & Norm
Masked Multi-Head Attention
Cheating!
<s>, Tu, es, un, sorcier, Harry, <eos>
Decoder
<s>, Tu, es, un, sorcier, Harry, <eos>

Decoder- Masking in self attention



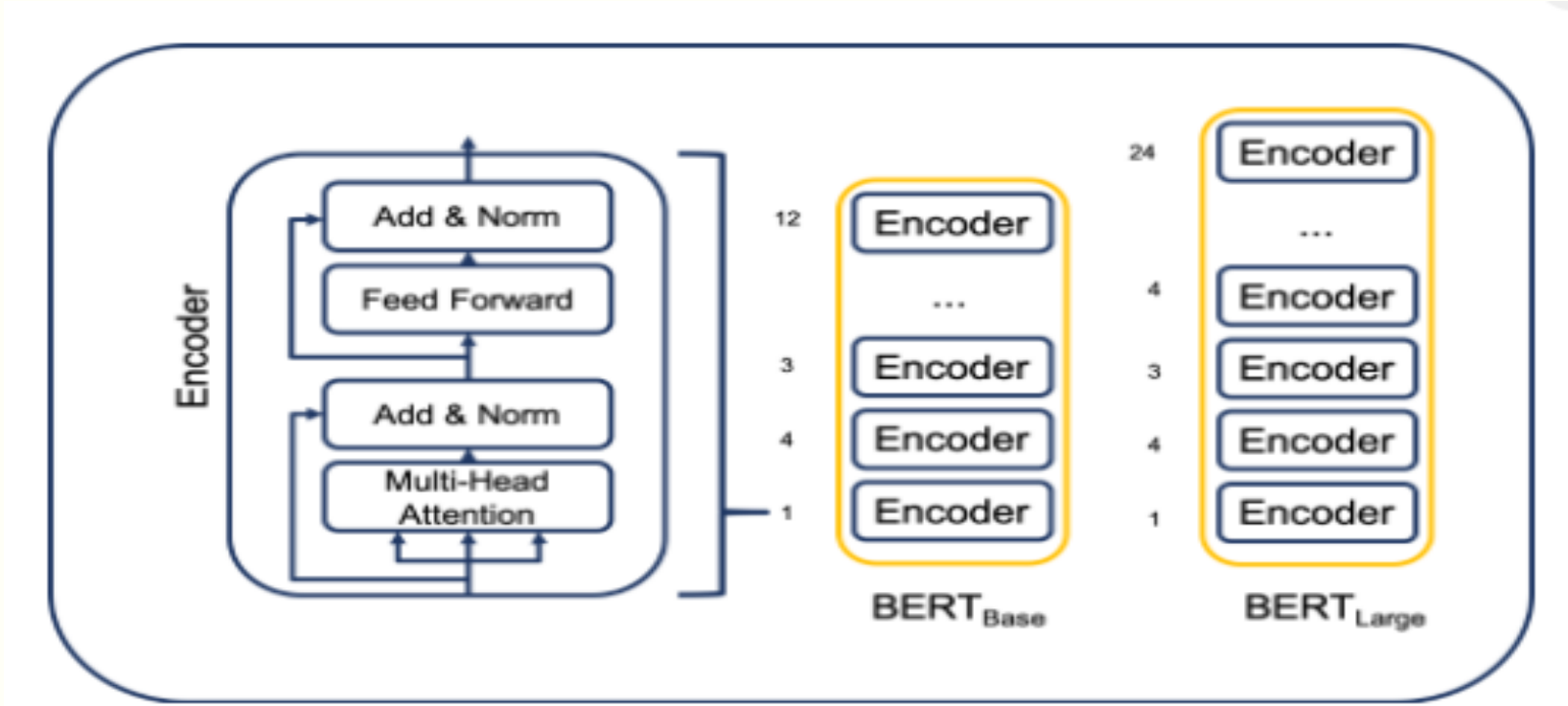
BERT Framework

- Bidirectional Encoder Representations from Transformers (BERT) is a **pre-trained** NLP model developed by Google.
- BERT is only an **encoder**, while the original transformer is composed of an encoder and decoder.
- BERT is a pre-trained deep learning **bidirectional model**.

BERT……

	Transformer Layers	Hidden Size	Attention Heads	Parameters	Processing	Length of Training
BERTbase	12	768	12	110M	4 TPUs	4 days
BERTlarge	24	1024	16	340M	16 TPUs	4 days

BERT…


Encoder
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
12
3
4
1
Encoder
...
Encoder
Encoder
Encoder
BERT Base
24
Encoder
...
4
Encoder
3
Encoder
4
Encoder
1
Encoder
BERT Large