



# Automated text summarization: state-of-the-art and prospects

Luca Cagliero  
Politecnico di Torino  
January 27 2023

# Natural Language Understanding



Politecnico  
di Torino  
SmartData@PoliTO

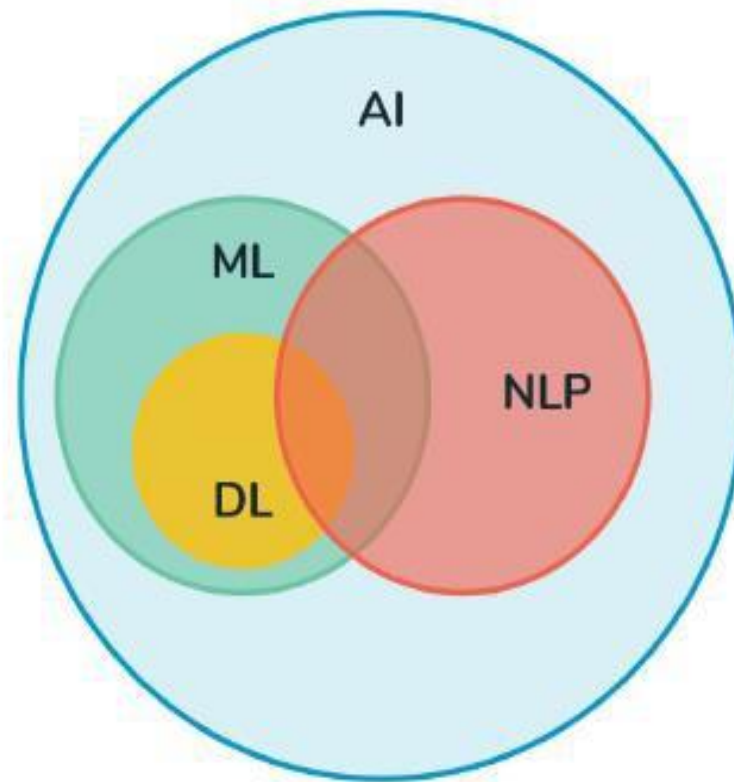


- NLU studies the connection between machine perception and languages
- Programming computers to fruitfully process large natural language corpora

# Natural Language Understanding

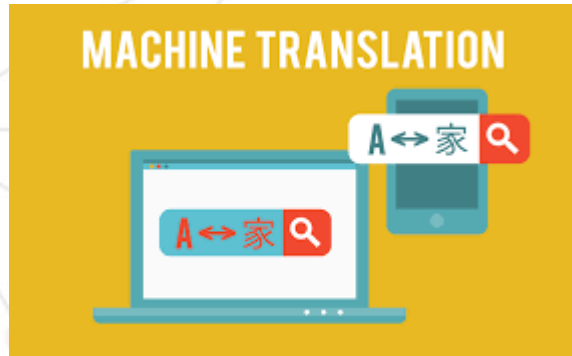


Politecnico  
di Torino  
SmartData@PoliTO



- Artificial intelligence
- Machine learning
- Language Processing
- Deep learning

# NLU applications



MACHINE TRANSLATION

"I am happy with this water bottle."



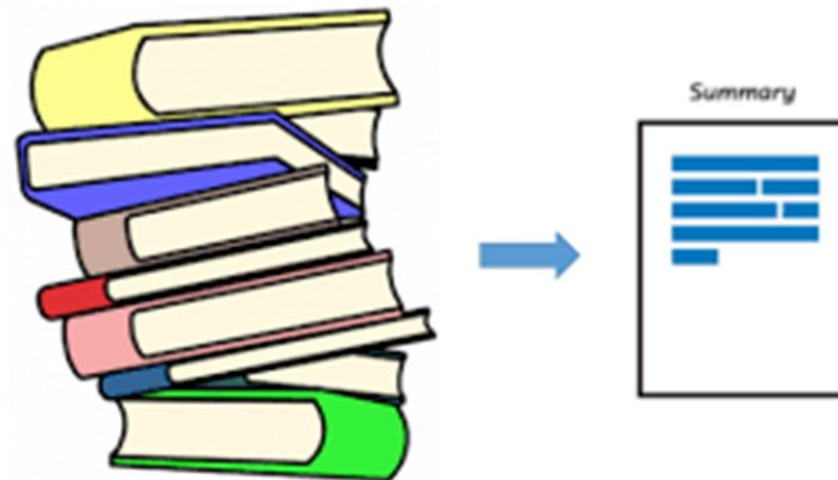
"This is a bad investment."



"I am going to walk today."



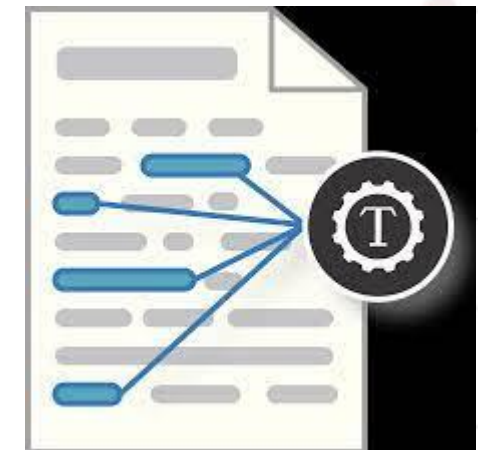
SENTIMENT ANALYSIS



TEXT SUMMARIZATION



AI CHATBOTS

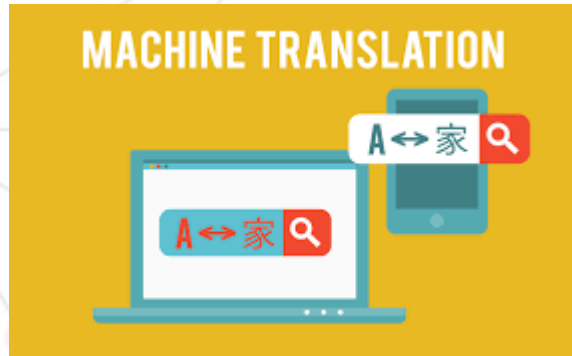


ENTITY RECOGNITION

# NLU applications



Politecnico  
di Torino  
SmartData@PoliTO



MACHINE TRANSLATION

"I am happy with this water bottle."



Positive

"This is a bad investment."



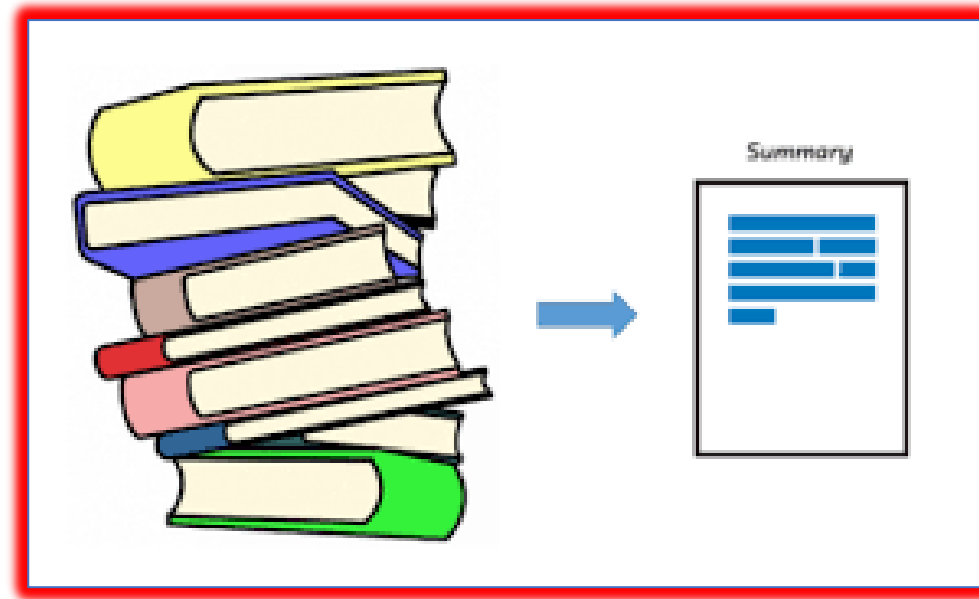
Negative

"I am going to walk today."



Neutral

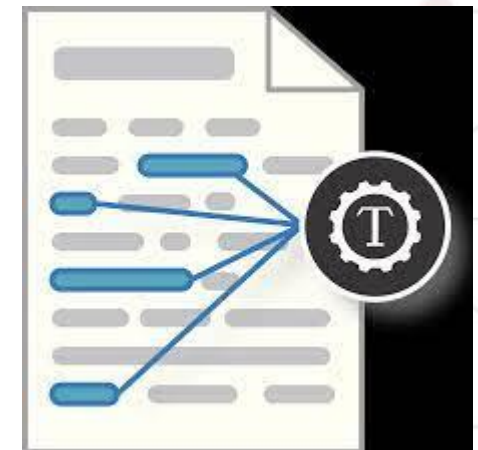
SENTIMENT ANALYSIS



TEXT SUMMARIZATION



AI CHATBOTS

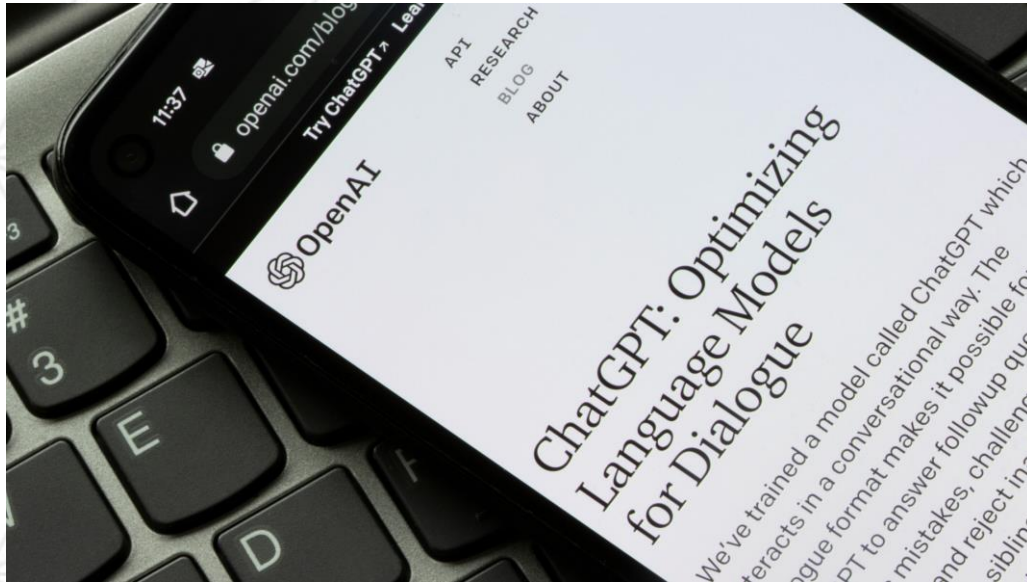


ENTITY RECOGNITION

# Impact of NLU



Politecnico  
di Torino  
SmartData@PoliTO



## THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Podcasts | Latest Headlines | More

Home World U.S. Politics Economy Business **Tech** Markets Opinion Books & Arts Real Estate Life & Work Style Sports

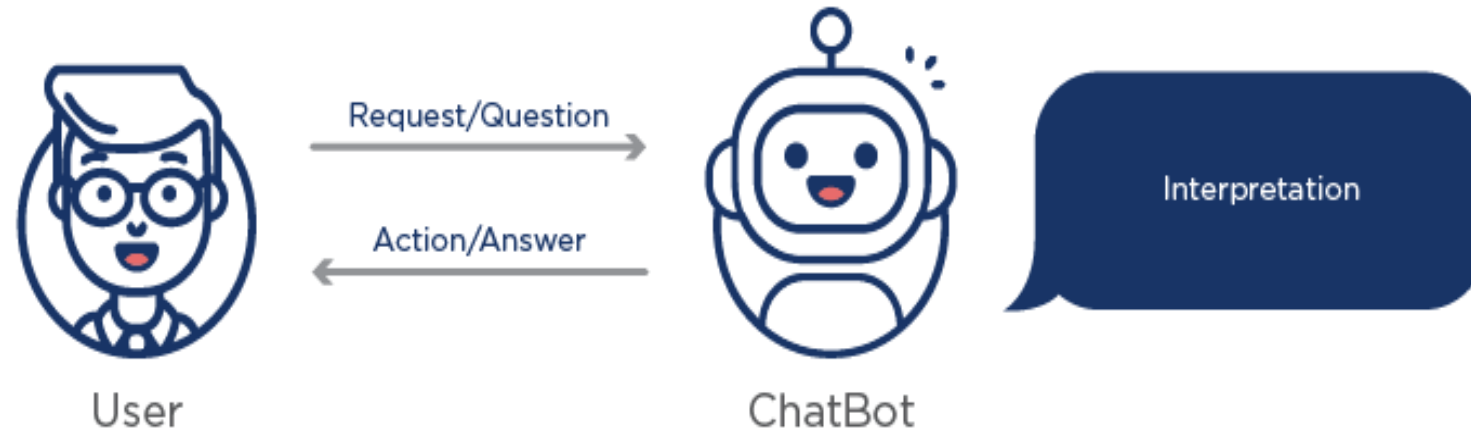
TECH

### Microsoft to Deepen OpenAI Partnership, Invest Billions in ChatGPT Creator

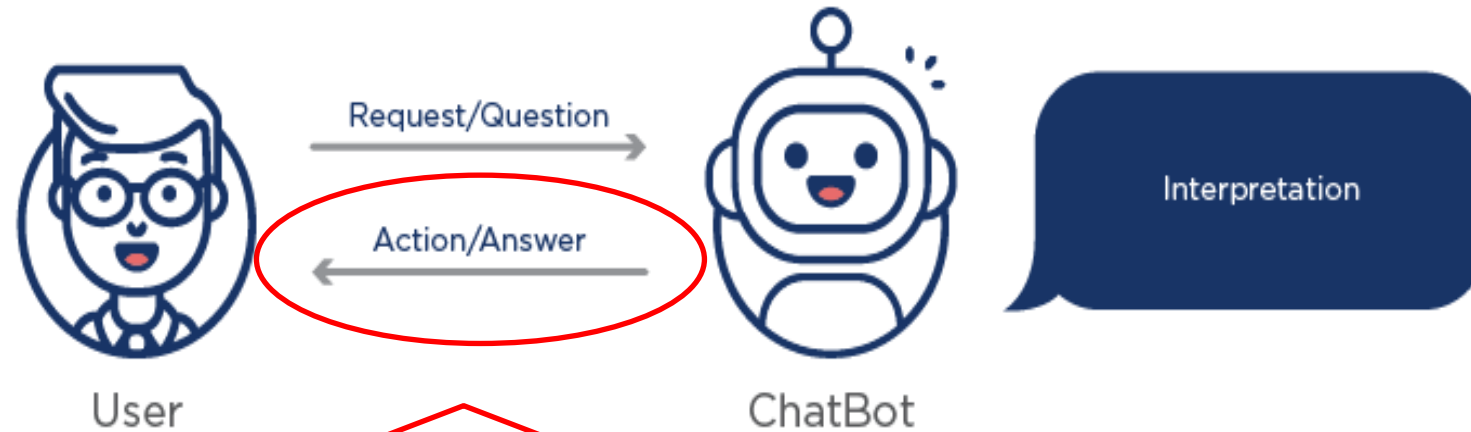
Software giant to expand use of artificial intelligence in its products



# Summarization in ChatGPT



# Summarization in ChatGPT



## USER EXPECTATION:

- Concise
- Informative
- Not redundant



# Text summarization in a nutshell



Politecnico  
di Torino  
SmartData@PoliTO



L

what's the goal of automated text summarization?



The goal of automated text summarization is to condense a large amount of text into a shorter, more concise version while still retaining the most important information and main ideas. This can be useful for tasks such as summarizing news articles, legal documents, or other lengthy texts in order to quickly understand their main points.



# Use cases



Politecnico  
di Torino  
SmartData@PoliTO




- Content curation
- Accessibility
- E-learning
- Business Intelligence
- ...



# Extractive vs. abstractive



L Please extract salient phrases from the paper "Attention is all you need"



- "self-attention mechanisms"
- "weigh the importance of each word"
- "handle longer sentences and larger input texts"
- "improve its performance"
- "parallel computing"
- "speed up the training process"
- "state-of-the-art results"
- "machine translation benchmarks"
- "widely used in other natural language processing tasks"
- "capture long-range dependencies"
- "handle large context"

Note that these are a few of the salient phrases in the paper, and that the paper covers many other details and experiments. These phrases are meant to give a general idea of the main contributions of the paper.

## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

# Extractive vs. abstractive



L

Please summarize the paper "Attention is all you need"



The paper "Attention is all you need" proposes a new model for neural machine translation called the Transformer. The Transformer model uses self-attention mechanisms to weigh the importance of each word in the input sentence when generating the output sentence. This allows the model to effectively handle longer sentences and larger input texts, improving its performance compared to previous models. The paper also uses parallel computing to speed up the training process, making it more efficient. The Transformer model achieved state-of-the-art results on several machine translation benchmarks and has since been widely used in other natural language processing tasks.



## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

# Cross-lingual summarization



Politecnico  
di Torino  
SmartData@PoliTO



L

Please explain me the goal of text summarization in French



Le but de la résumé automatisé de texte est de condenser un grand texte en une version plus courte et plus concise, tout en conservant les informations les plus importantes et les idées principales. Cela peut être utile pour des tâches telles que le résumé d'articles de presse, de documents juridiques ou d'autres textes longs afin de comprendre rapidement leurs points principaux.



## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.



# Query-driven summarization



Politecnico  
di Torino  
SmartData@PoliTO



L Please summarize now the content of the paper "Attention is all you need" to a AI expert



The paper "Attention is all you need" presents a new architecture for neural machine translation, the Transformer, which utilizes self-attention mechanisms to weigh the importance of words in input sentences when generating the output sentence. This allows the model to handle longer sentences and larger input texts more effectively, resulting in improved performance compared to previous models. The authors also introduce parallel computing to speed up the training process. The Transformer model achieved state-of-the-art results on several machine translation benchmarks and has since been widely adopted in other NLP tasks due to its ability to capture long-range dependencies and handle large context.



L Please summarize the paper "Attention is all you need"



The paper "Attention is all you need" proposes a new model for neural machine translation called the Transformer. The Transformer model uses self-attention mechanisms to weigh the importance of each word in the input sentence when generating the output sentence. This allows the model to effectively handle longer sentences and larger input texts, improving its performance compared to previous models. The paper also uses parallel computing to speed up the training process, making it more efficient. The Transformer model achieved state-of-the-art results on several machine translation benchmarks and has since been widely used in other natural language processing tasks.



## Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.



# Aspect-based summarization



1<sup>st</sup> ASPECT

**L** Describe the evaluation process of a text summarization from both qualitative and quantitative perspectives



Evaluating the performance of text summarization models can be done from both a qualitative and a quantitative perspective.



From a qualitative perspective, the summary can be evaluated by a human evaluator for its grammatical correctness, fluency, relevance, and coherence. The evaluator can also check if the summary captures the main ideas of the original text and if it is able to convey the meaning of the text in a shorter form.

2<sup>nd</sup> ASPECT

From a quantitative perspective, there are several metrics that can be used to evaluate the performance of text summarization models. Some of the most commonly used metrics include:

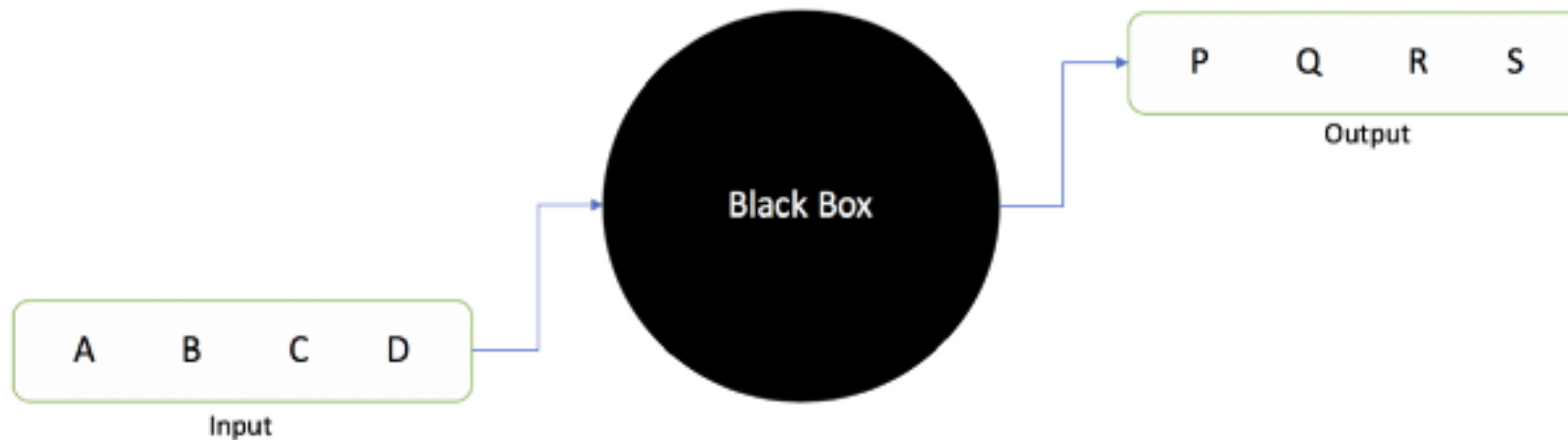
- ROUGE: Recall-Oriented Understudy for Gisting Evaluation, which compares the n-grams (word sequences) of the summary and the reference summary, and calculates a score based on the number of overlapping n-grams.
- METEOR: METric for Evaluation of Translation with Explicit ORdering, which is a modified version of ROUGE that also takes into account synonyms and paraphrases.
- BLEU: Bilingual Evaluation Understudy, which is similar to ROUGE, but also takes into account the order of the words in the summary and reference summary.
- CIDEr: Consensus-Based Image Description Evaluation, which is a more advanced metric that takes into account not only the n-grams but also the semantics of the summary and reference summary.

The most common approach is to use a combination of both qualitative and quantitative evaluation to get a comprehensive understanding of the model performance.

# State-of-the-art approaches



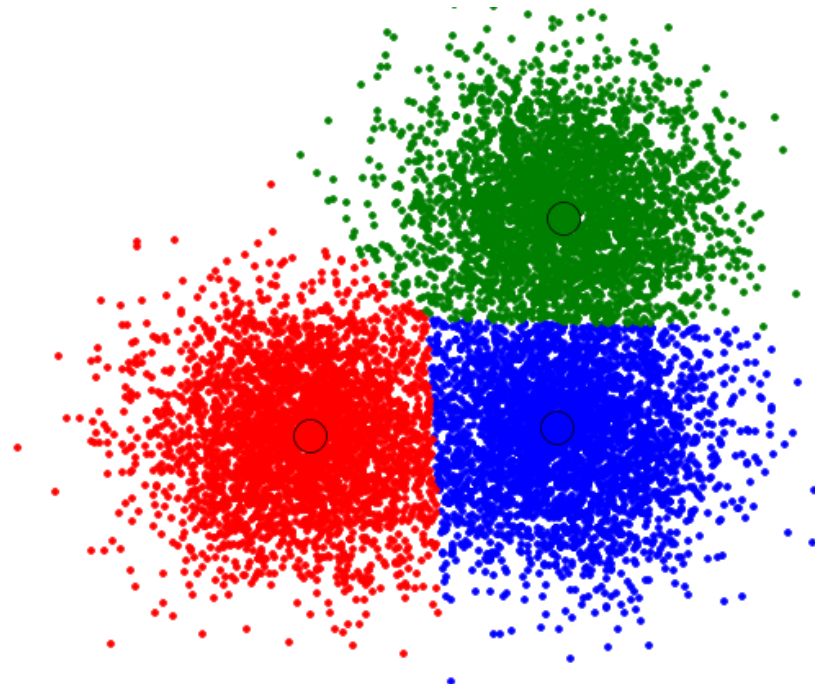
- Based on Deep Learning
- Leverage Sequence-to-Sequence Models
  - As many other NLU applications
    - E.g., Machine Translation, Question Answering, Text Style Transfer, etc.



# One step back...



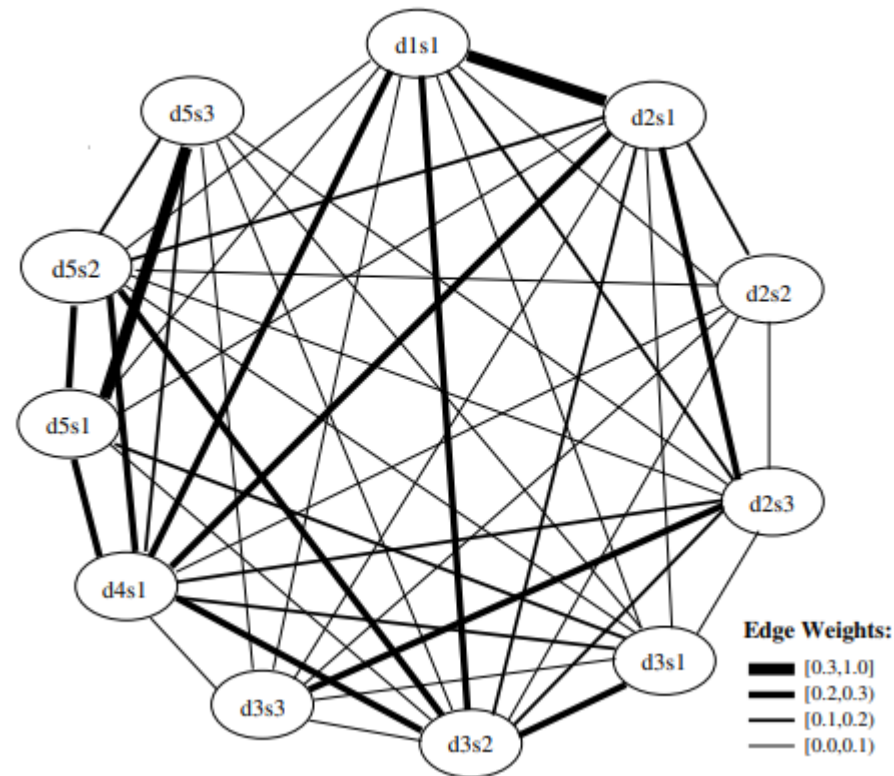
- Before the advent of Deep Learning...
- **Unsupervised methods** based on
  - **Clustering**
  - Graphs
  - Pattern mining
  - Integer Linear Programming
  - ...



# One step back...



- Before the advent of Deep Learning...
- **Unsupervised methods** based on
  - Clustering
  - **Graph mining**
  - Pattern mining
  - Integer Linear Programming
  - ...



# One step back...



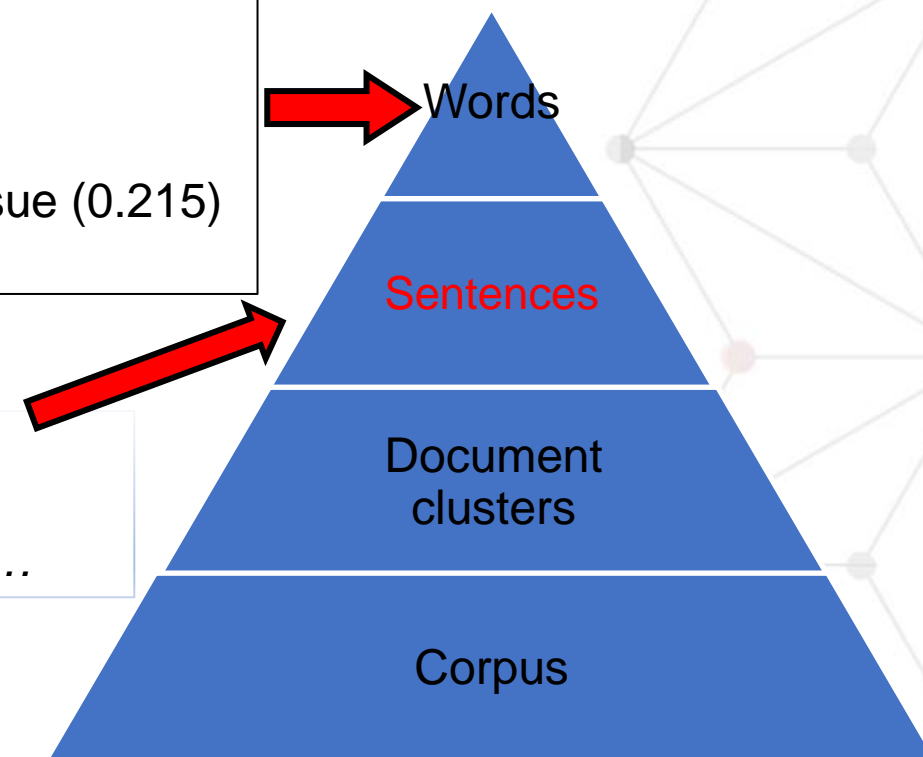
- Before the advent of Deep Learning...

- **Unsupervised methods** based on

- Clustering
- Graph mining
- **Pattern mining**
- Integer Linear Programming
- ...

Gener Diseases (0.778)  
Diseas Disorder (0.774)  
Stem Cell (0.671)  
...  
Stem Cell Esophag Tissue (0.215)  
...

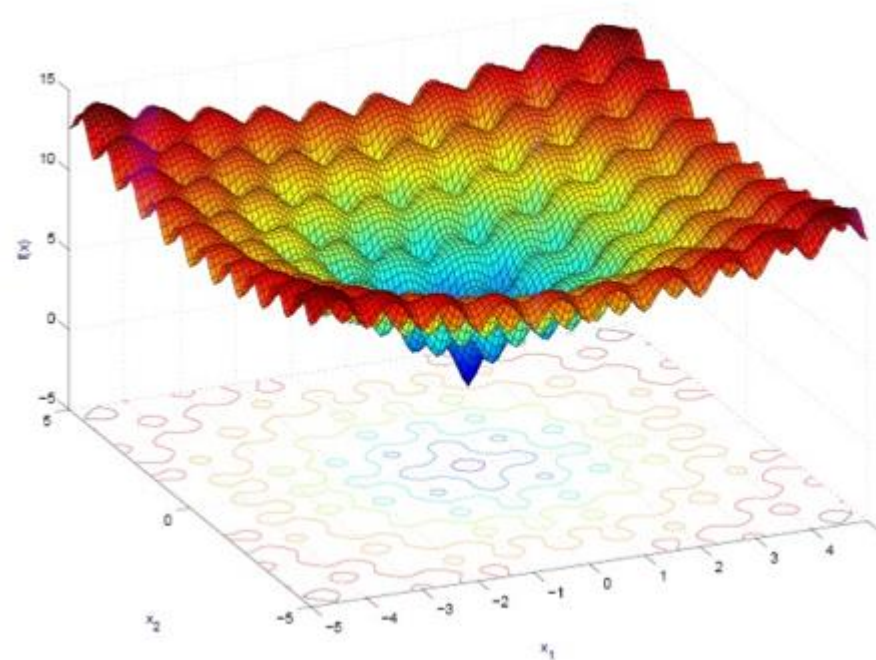
*This treatment for a disease ...*  
~~*This is another treatment ...*~~  
*This work covers different topics ...*



# One step back...



- Before the advent of Deep Learning...
- **Unsupervised methods** based on
  - Clustering
  - Graph mining
  - Pattern mining
  - **Integer Linear Programming**
  - ...

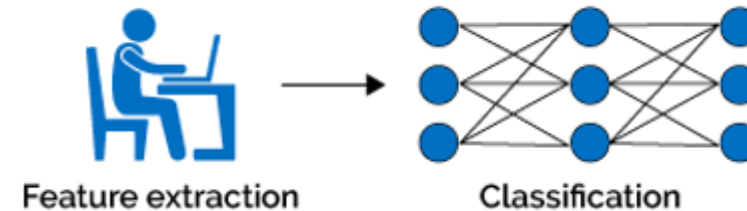
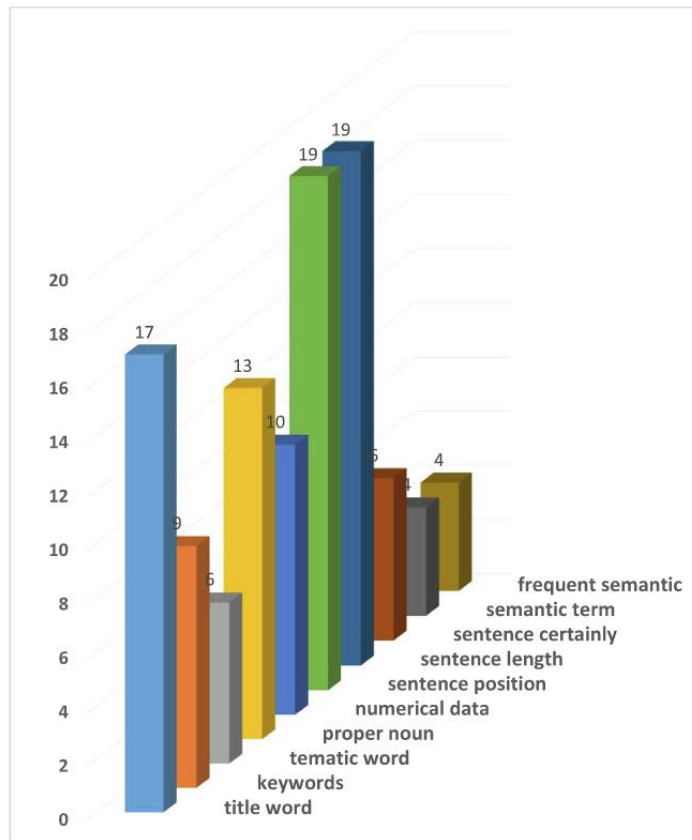




# One step back...



- Before the advent of Deep Learning...
- **Supervised methods** based on selected features



# In the present day...



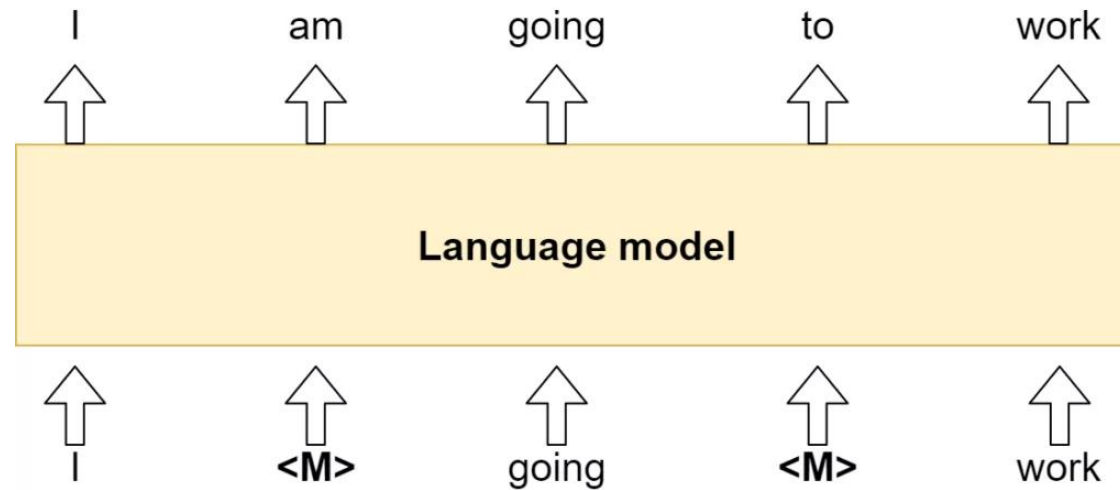
- Inference of **semantic text relations** by leveraging the inherent text structure



# In the present day...



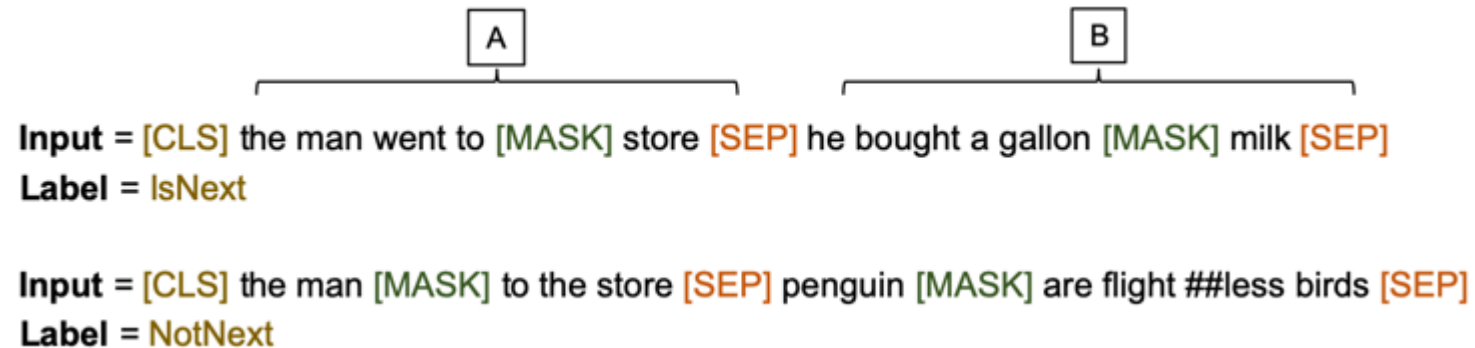
- From **supervised** to **self-supervised**
- **BERT pretraining**



# In the present day...



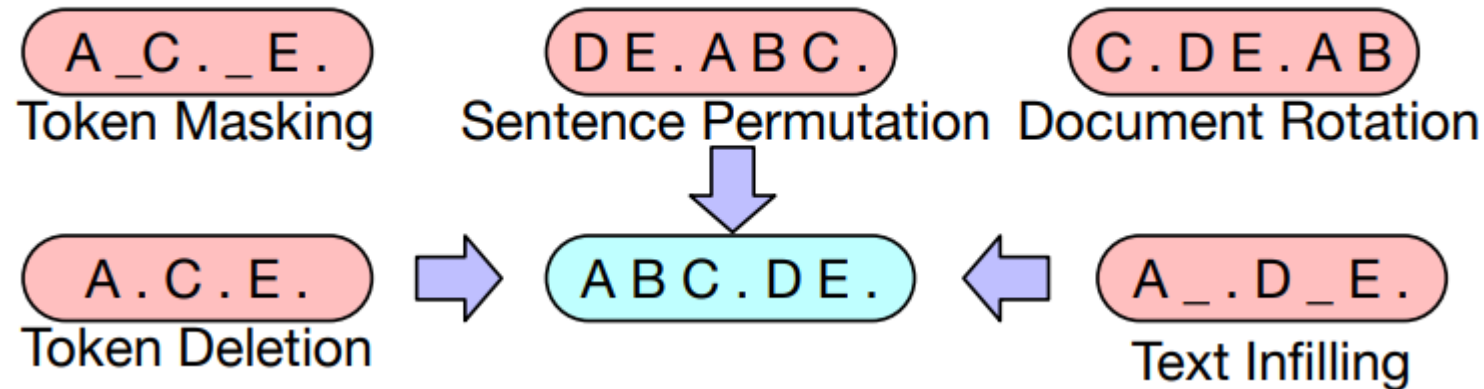
- From **supervised** to **self-supervised**
- **BERT** pretraining



# In the present day...



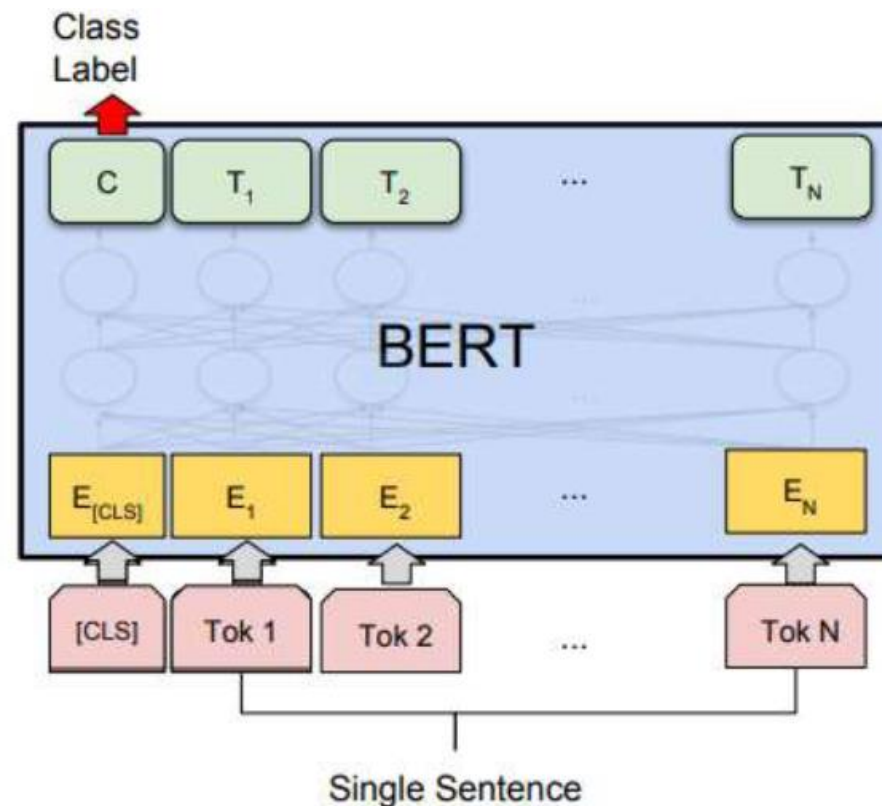
- From **supervised** to **self-supervised**
- **BART** pretraining



# In the present day...



- **Fine-tuning** of the pretrained model
  - **Sentence classification**
- Given a sentence, BERT returns the most likely class label
  - **Label=True** means that the sentence belongs to the summary
  - **Label=False** otherwise

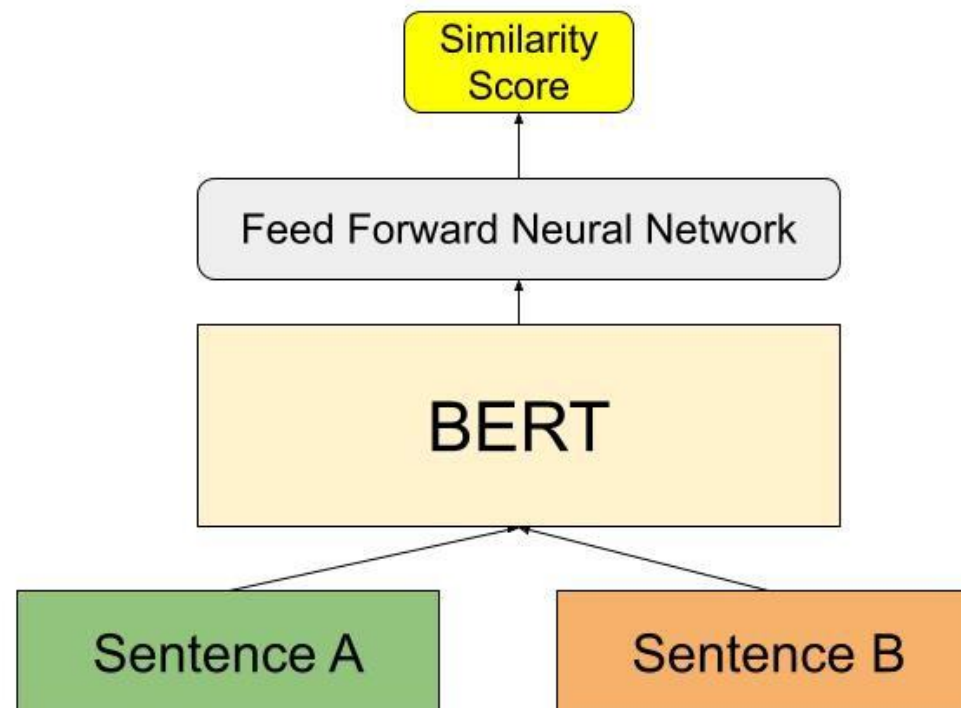




# In the present day...



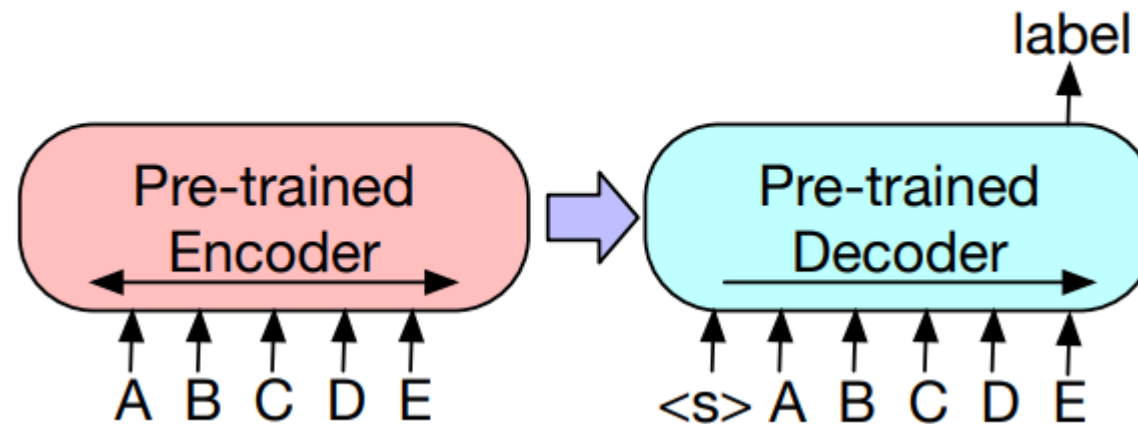
- **Fine-tuning** of the pretrained model
  - **Sentence similarity**
    - Given a pair of sentences, BERT returns the similarity score



# Abstractive summarization



- Sequence-to-sequence model for **abstractive text summarization**
- Now also in **Italian!**
  - La Quatra & Cagliero, BART-IT: Italian pretraining for BART sequence to sequence model. 2023. <https://doi.org/10.3390/fi15010015>
  - Available at <https://huggingface.co/spaces/moreno/q/italian-summarization>





# Thanks for the attention!



L

summarizing data always possible?



Summarizing data is not always possible, depending on the complexity and characteristics of the data. Some types of data, such as images or videos, might not be easily summarized, and a more detailed analysis might be necessary.  

Additionally, the effectiveness of summarization also depends on the specific task and the desired level of information retention. For example, in some cases, a summary that retains only the most important or salient information might be sufficient, while in other cases, a more detailed summary might be