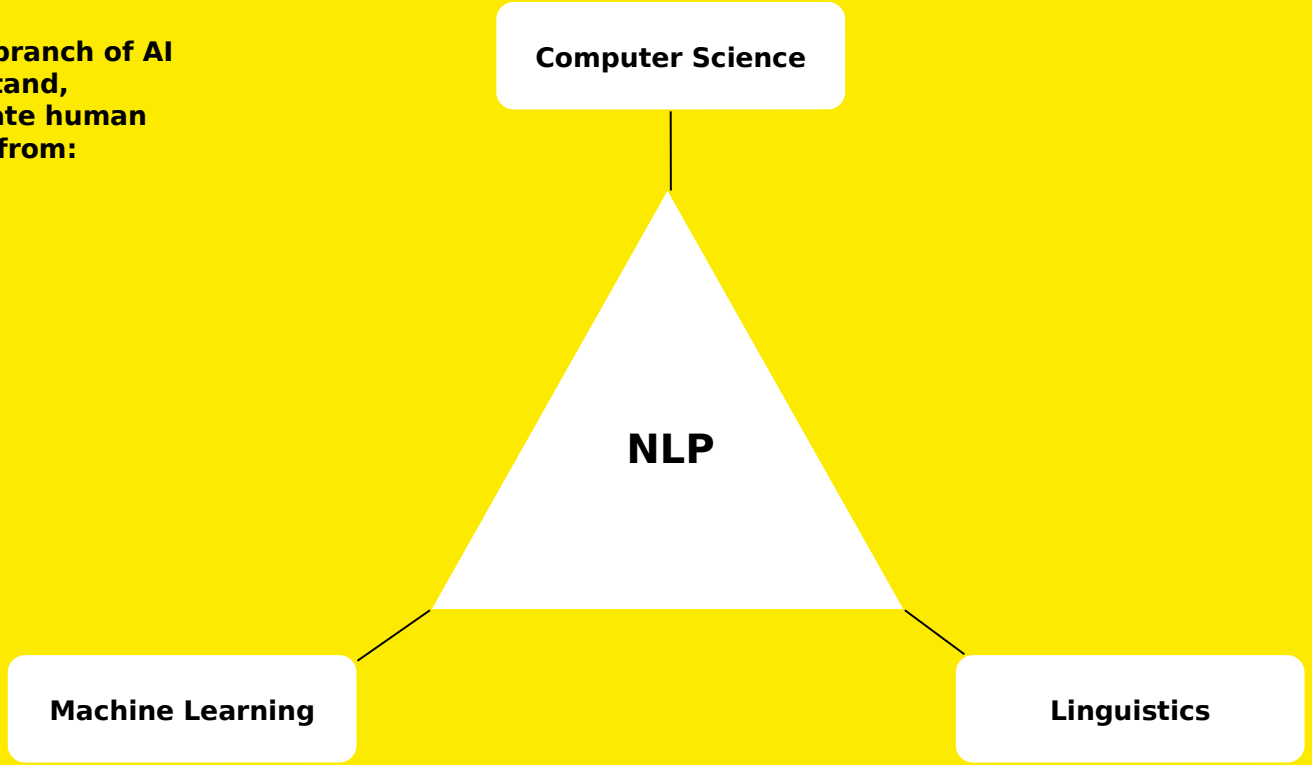
Aidan Klug and Muhammad Usman Nawaz

Natural Language Processing

What is Natural Language Processing?

Natural Language Processing is a branch of AI that enables computers to understand, interpret and subsequently generate human language. It combines techniques from:

- **Computer science**
- **Machine Learning**
- **Linguistics**



Why Should we care about NLP?

- NLP powers the apps we use every day:
- 📱 Virtual assistants like Siri & Alexa
- ✉️ Spam filters in Gmail
- 🌐 Google Translate
- 💬 ChatGPT and customer service chatbots
- It bridges the gap between **human language** and **machine understanding**
- The better NLP gets → the more natural and helpful AI becomes

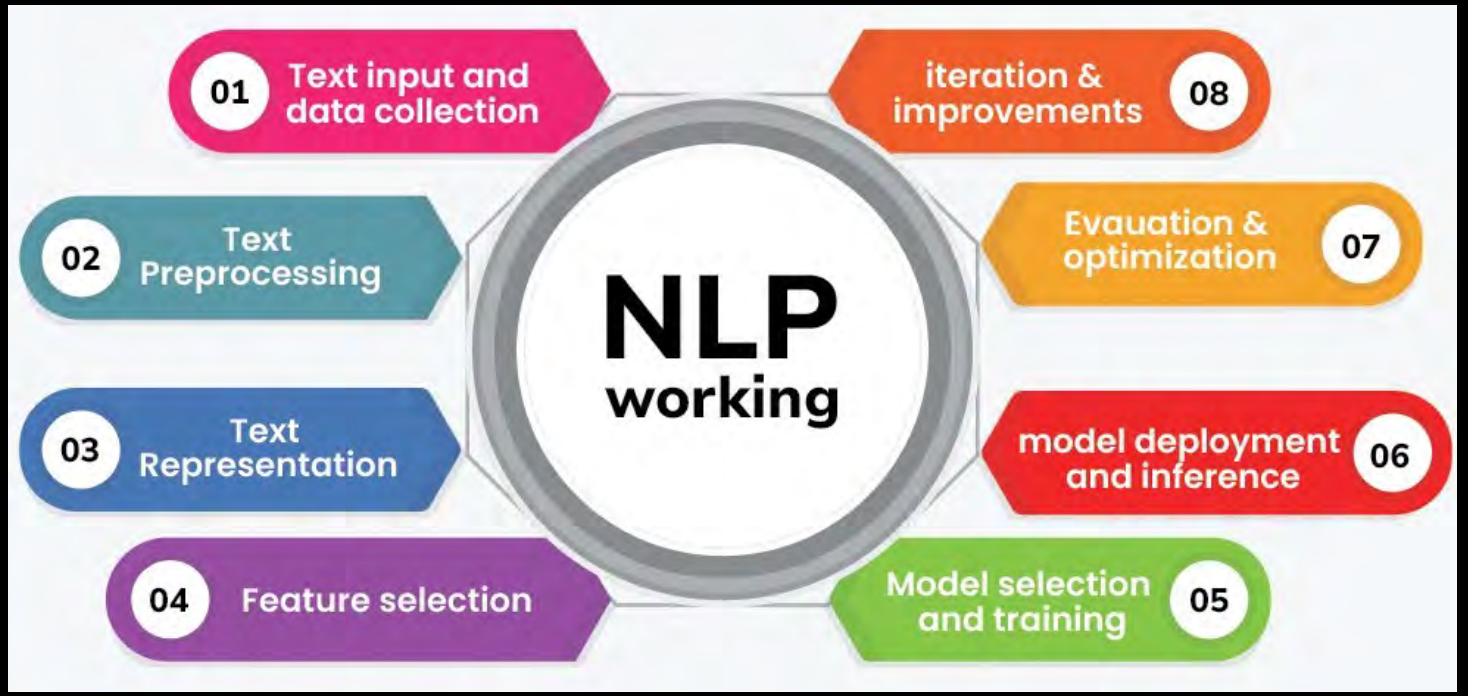
Importance of NLP

- Data Abundance: A significant portion of data is unstructured text; NLP helps in meaningful information.
- Automation: Facilitates automation in customer service, content moderation and more.
- Accessibility: Enhances accessibility through speech recognition and translation services.

Real-World Examples of NLP

Application	NLP Task
Gmail Auto-Complete	Language Modeling
ChatGPT	Text generation
Google Translate	Machine Translation
Grammarly	Grammar Correction
Amazon Alexa	Speech Recognition & Intent Detection

What is Natural Language Processing (NLP) Used For?

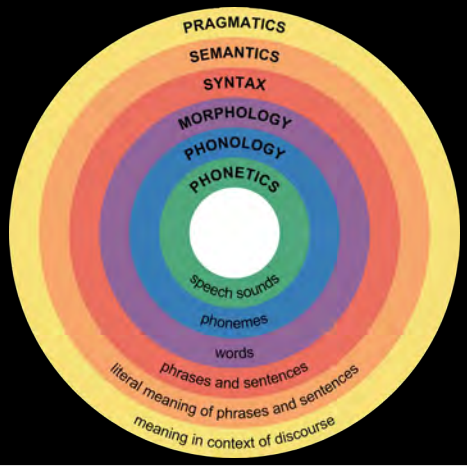


NLP HISTORY

This Guy

In the early 1900's, Swiss linguistics professor (and legendary mustache proprietor) **Ferdinand De Saussure** developed an approach to describing languages as systems.

- Each sounds represents a concept
- That concept shifts meaning as the context changes

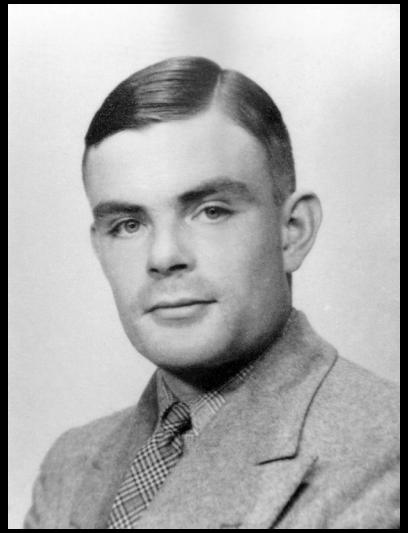


Died in 1913 before publishing his theories but his colleagues published his works in 1916

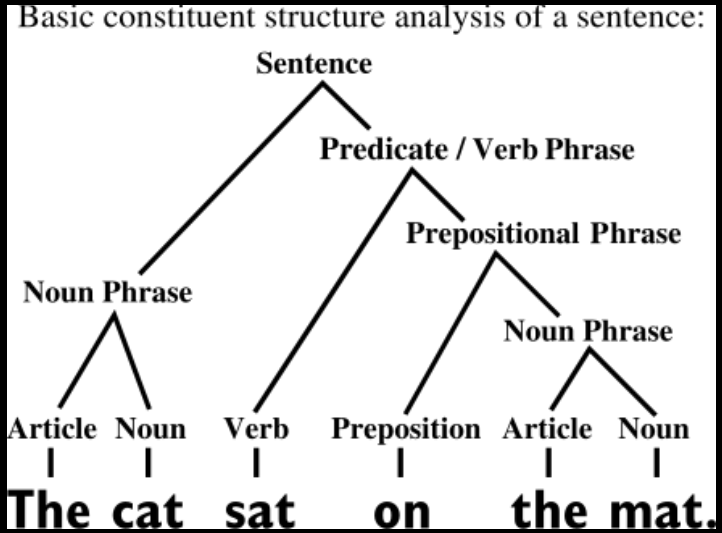
- This laid the groundwork for the structuralist approach to language classification
- This theory emphasizes relationships between words, their meaning and the ways they evolve in a sentence or paragraph.

NLP HISTORY

Later in 1950, Alan Turing wrote a paper that described a set of parameters to test a thinking machine that would be capable of conversation through the use of a teleprinter



In 1957, Noam Chomsky published Syntactic Structures, creating a style of grammar called phase structure grammar, which translated sentences into a format that is usable by computers



Core NLP Tasks

- Tokenization
- Part of Speech Tagging
- Name Entity Recognition
- Sentiment Analysis
- Machine Translation

NLP APPROACHES

Rules Based

Earliest NLP used simple decision trees with preprogrammed rules - NOT ML!

Statistical

Automatically extracts, classifies and labels the elements of in text or voice data, then applies a statistical likelihood to the meaning of those elements

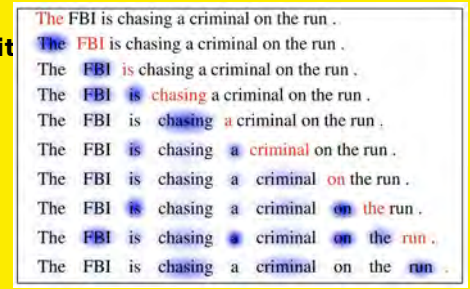
Deep Learning

Convolutional neural networks, recurrent neural networks, auto-encoders and transformers are all examples of more modern NLP techniques that leverage deep learning

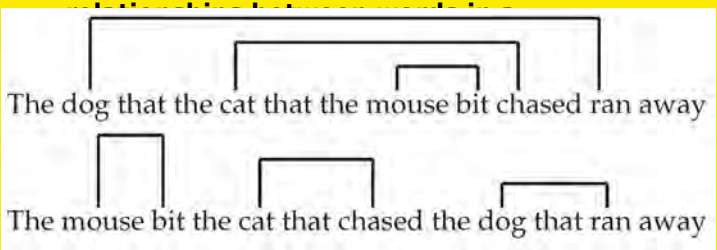
Transformers

Transformers are an NLP approach that utilize neural network architecture to process text. They support:

- **Parallel Processing: allowing them to multiple words in a sentence at the same time. This makes them more efficient and faster compared to other techniques**
- **Self-Attention: Mechanisms within the structure of the transformer allows it to weigh the importance of different words within the sentence**



- **Long-Range Dependency: Transformers are able to pick out**



Encoders VS Decoders

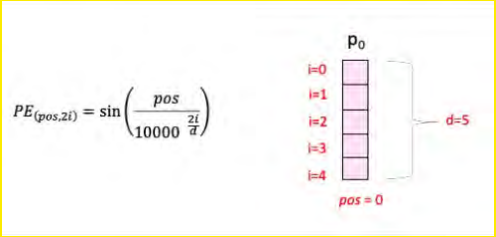
- **The encoder takes input data and transforms it into an embedding that captures the overall meaning. This meaning is summarized as a vector with both attention information (how much emphasis each word has in the sentence) and positional information (where that word occurs in the sentence).**

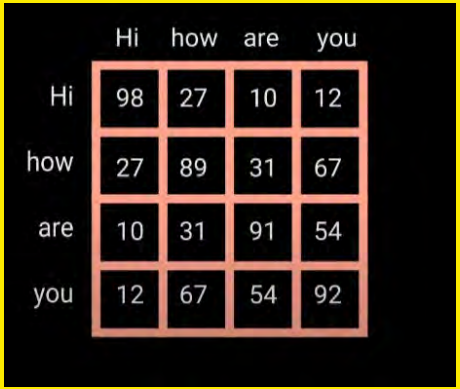
- **Decoder generates output back to user one word at a time by recursively finding the next most probable word in the sentence. The input to this recursive process is both the original encoder information AND the previous output from the decoder layer.**

Encoders VS Decoders

Encoder Steps:

1. **Input Embedding - A lookup table to grab a learned vector representation of each word in the input**
2. **Positional Encoding - Inject position information about each word into the vector so as to keep track of each words position in the overall sentence**
3. **Encoder Layer**
   a. **Multi-headed attention - Allows model to associate each word in the input to others. Makes output vector.**
   b. **Feed Forward - Multi-head vector is added to the original input vector (residual connection). This undergoes layer normalization and is then repeatedly fed into the feed forward network. This process allows a reduction in training time as well as a refinement of the attention vector**

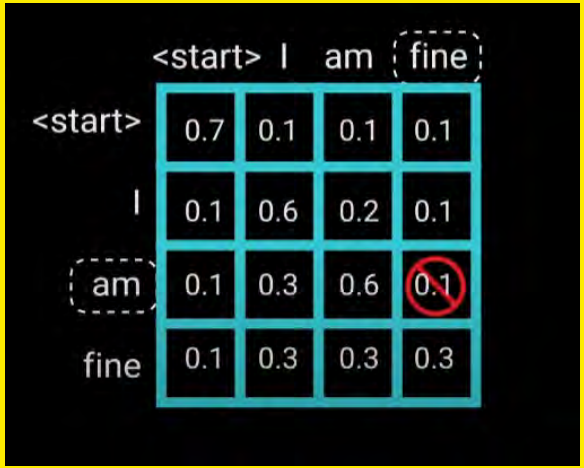




Encoders VS Decoders

Decoder Steps:

1. **Takes the encoders output and attention information.**
2. **Generates the first word in the response.**
3. **Continually takes the encoders output and attention information AND the previously generated decoder output, to predict the next word.**
4. **A multi-headed attention layer acts on the decoders output and prevents the decoder from having accounting for future words. This is called masking.**
5. **Repeats this process word by word until the network generates an end token.**



EXAMPLES

- **BERT - Bidirectional encoder representations from transformers. Can work forwards and backwards on sentences.**
- **GPT - Unidirectional autoregressive decoder model.**
- **ELMO - Embeddings from Language models, contextual word embedding model. Bidirectional like bert and uses long short-term memory (LSTM) to solve issues associated with vanishing gradients present is some other techniques.**
- **UniLM - Unified pre-trained langauge model - an amalgamation of uni / bi - directional and sequence to sequence prediction.**

Thank You

Recommended Resources

Transformers (how LLMs work) explained visually | DL5 - 3Blue1Brown

IBM Crash Course - https://www.ibm.com/think/topics/natural-language-processing
https://www.deeplearning.ai/resources/natural-language-processing/