**What is NLP?**

NLP stands for Natural Language Processing. It is the branch of AI. It gives the machines an ability to understand the human language and process it. Human Language can be in the form of text or audio format. In short it programs computers to process large amount of natural language data

Eg: Contextual Advertisements, Spam filtering, Removing Adult Content, Opinion mining, Search Engines(Google), Chatbots

While learning new things, I believe we need to start with the history( where it started, how it started and the future importance of it) and the challenges faced . If we get to know the history then we will be knowing the value

**History of NLP:**

**NLP started in 1950 when Alan Mathison Turing published an article in the name Computing Machinery and Intelligence. It talks about automatic interpretation and generation of natural language. Different approaches have come to deal with NLP tasks**

* **Heuristic-Based NLP:**

**A traditional and quick approach that relies on predefined set of rules and patterns to process Natural Language( regex, wordnet - a lexical dictionary, open mind common sense)**

**Eg: Early Spam filters use Heuristics to find spam related keyword which are commonly used in finding the spam emails**

**Another example would be finding proper nouns based on capitalization of patterns in English text**

**When there are too many rules and the problem is open ended ml based nlp comes into play**

* **Statistical Machine Learning based NLP:**

**It is based on statistical rules and machine learning algorithms. In this approach algorithms are applied to the data and learned from the data and applied to various tasks**

**Eg: Naive Bayes, Support Vector Machine, Hidden Markov Model(HMM), Logistic Regression, LDA**

**When text is converted into numbers. Most of the times sequential information which is present inside the text (sentence formation from left to right) they doesn't care about it then deep learning based NLP comes into play**

**Feature generation is automatically done by deep learning**

* **Neural Network based NLP/ Deep Learning based NLP:**

**This is the latest approach and involves neural network based learning which is also known as deep learning. This approach is very data hungry, time consuming and requires high computational power.**

**Eg: RNN, CNN, LSTM**

**Challenges in NLP:**

**Ambiguity :** There is one or more meanings for some sentences

**Contextual words: eg:**I ran to the store because we ran out of milk

**Slang and Colloquialisms:** “Piece of cake”, Pulling your leg

**Synonyms**: words which has same meaning

**Irony, Sarcasm/Tonal difference**: eg: That’s just what i needed today

**Spelling errors**

**Creativity**

**Diversity**

**Components of NLP:**

**NLP consists of two components:**

1. **Natural Language Understanding**
2. **Natural Language Generation**

**Eg of NLP:**

**Voice Assistants like Google Alexa and Siri because Text and Speech processing is involved**

**Text classification apps like Google docs, Microsoft Word, Grammarly**

**Phases of NLP:**

**Lexical Analysis:**

This is the first phase of NLP and it converts lines into sequence of tokens( A meaningful unit such as punctuation mark, word, number)

Eg : int x = 5; Lexemes are ‘int’, ‘x’, ‘=’, ’5’, ’;’

**Syntactic Analysis:**

This is the process of analyzing seq of tokens grammar language.

Eg:

The cat sat on the mat.

* The parser might produce a tree with "sat" as the root verb, "The cat" as the subject, and "on the mat" as a prepositional phrase modifying the verb.

Two types of approaches are bottom up and top down parsing

**Semantic Analysis:**

It involves understanding the meaning of a given piece of text or code after it has been syntactically analyzed. The main goal is to ensure the structure identified during syntactic analysis makes sense logically and contextually.

Eg:

Int x = 5;

Float y = “hello”; //Error

During Semantic Analysis ‘float y = “hello”’ will raise an error because ‘hello’ is a string, not a float, and this assignment violates type system of C

**Pragmatic Analysis:**

It seeks to comprehend how language is used in specific situations, taking into account the intentions, beliefs and the context of the speaker or writer, as well as how the language might be interpreted by the listener or reader.

Eg: Indirect Requests

Sentence: “Could you pass the Salt”?

Although it is framed as a question, it’s understood as polite request

**Disclosure Integration:**

Process of combining disclosed info from multiple sources into a coherent or unified representation

Eg: Generating a summary that integrates disclosures from different sections of a legal contract, such as risks, obligations, and compliance requirements, into a single summary document.

**Common NLP Tasks:**

1. **Text/Document Classification**
2. **Sentiment Analysis**
3. **Information Retrieval**
4. **Parts of Speech Tagging**
5. **Language Detection/ Machine Translation**
6. **Conversational Agents**
7. **Knowledge Graph and QA systems**
8. **Text Summarization**
9. **Topic Modeling**
10. **Text Generation**
11. **Spell checking and grammar correction**
12. **Text Parsing**
13. **Speech to text**