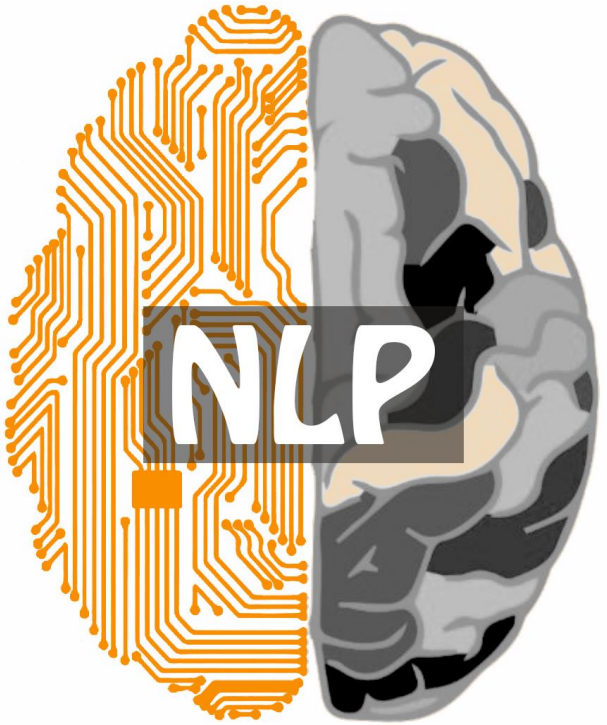
INTRODUCTION TO NLP

Dr. Krishnendu Guha
Assistant Professor (On Contract)
National Institute of Technology (NIT), Jamshedpur
Email: krishnendu.ca@nitjsr.ac.in

INTRODUCTION

- **Natural Language Processing**, usually shortened as NLP,
- is a branch of artificial intelligence that deals with the interaction between computers and humans using the **natural language**



- The ultimate objective of NLP is
- to read,
- decipher,
- understand, and
- make sense of the human languages in a manner that is valuable

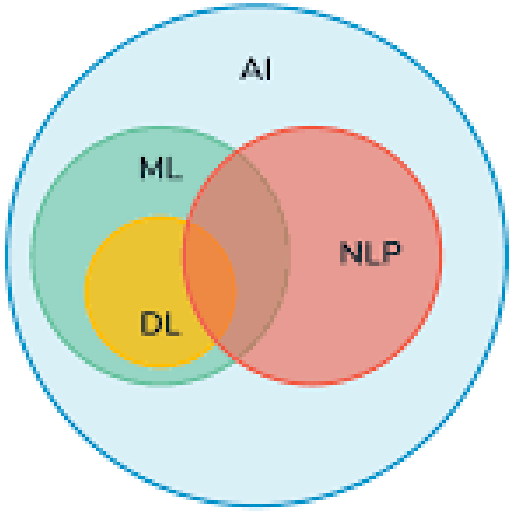
- Most NLP techniques rely on machine learning to derive meaning from human languages

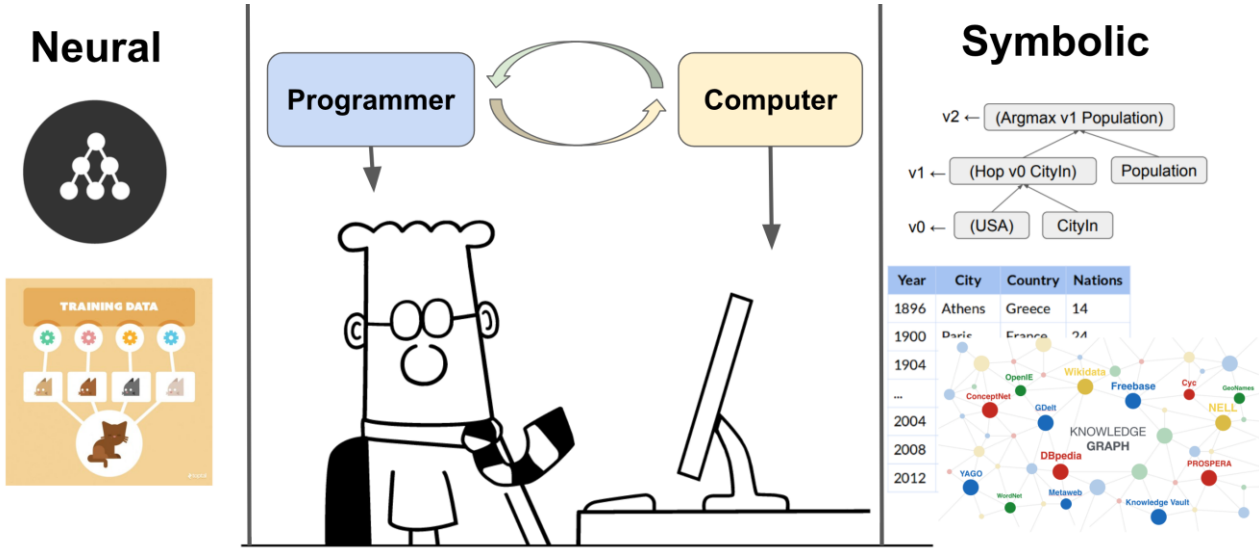
- ***OR SIMPLY***
- Natural Language Processing is the technology used to aid computers to understand the human's natural language

- It's not an easy task teaching machines to understand how we communicate

- **HISTORY**

- Symbolic NLP (1950s - early 1990s)
  - collection of rules, with questions and matching answers
- Statistical NLP (1990s - 2010s)
  - based on complex sets of hand-written rules
- Neural NLP (present)
- representation learning and deep neural network-style machine learning methods





HUMAN MACHINE INTERACTION USING NLP

- A typical interaction between humans and machines using Natural Language Processing could go as follows:
- 1. A human talks to the machine
- 2. The machine captures the audio
- 3. Audio to text conversion takes place
- 4. Processing of the text's data
- 5. Data to audio conversion takes place
- 6. The machine responds to the human by playing the audio file

WHAT IS NLP USED FOR?

- Natural Language Processing is the driving force behind the following common applications:
- Language translation applications such as Google Translate

- Word Processors such as Microsoft Word and Grammarly that employ NLP to check grammatical accuracy of texts.

- Interactive Voice Response (IVR) applications used in call centers to respond to certain users' requests.

- Personal assistant applications such as OK Google, Siri, Cortana, and Alexa.

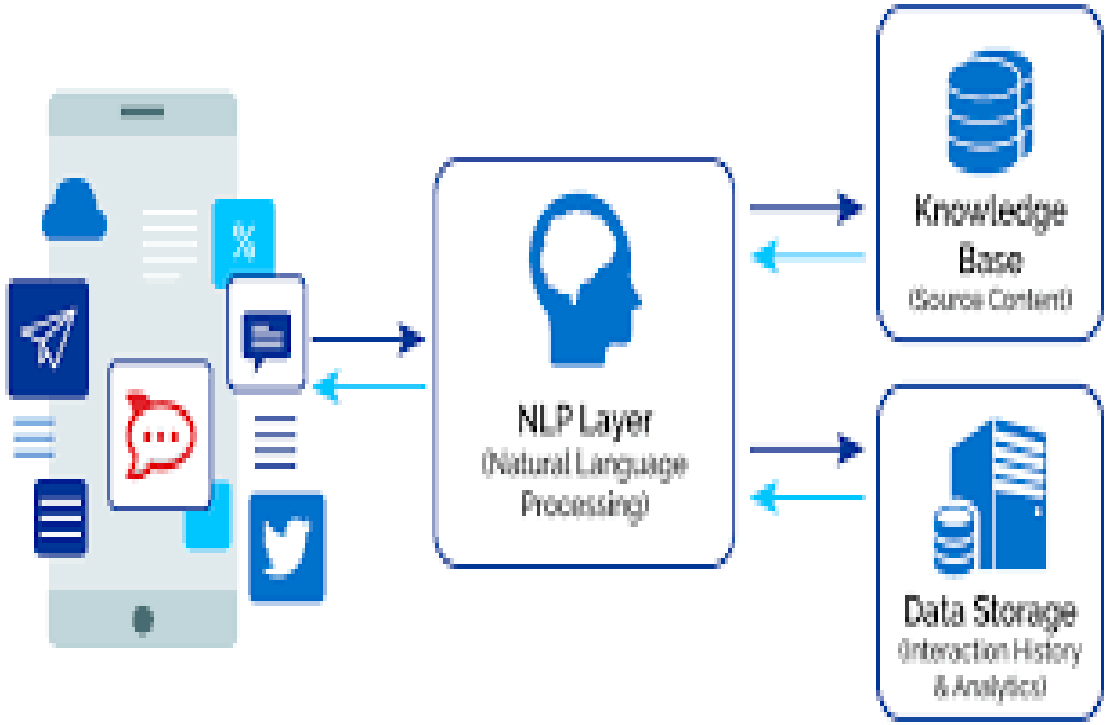
WHY IS NLP DIFFICULT?

- It's the **nature of the human language** that makes NLP difficult.
- The **rules that dictate the passing of information** using natural languages are not easy for computers to understand.
- Some of these **rules can be high-leveled and abstract**;
  - for example, when someone uses a sarcastic remark to pass information.
- On the other hand, some of these **rules can be low-levelled**;
  - for example, using the character "s" to signify the plurality of items.
- Comprehensively **understanding the human language requires understanding both the words and how the concepts are connected to deliver the intended message**.
- While humans can easily master a language, the **ambiguity and imprecise characteristics** of the natural languages are what make NLP difficult for machines to implement.

HOW NLP WORKS?

- NLP entails **applying algorithms to identify and extract the natural language rules** such that the unstructured language data is converted into a form that computers can understand
- When the text has been provided, the computer **will utilize algorithms to extract meaning associated with every sentence and collect the essential data** from them.

- Sometimes, the computer may fail to understand the meaning of a sentence well, leading to obscure results.
- **For example**, a humorous incident occurred in the 1950s during the translation of some words between the English and the Russian languages.
- Here is the biblical sentence that required translation:
- *"The spirit is willing, but the flesh is weak."*
- Here is the result when the sentence was translated to Russian and back to English:
- *"The vodka is good, but the meat is rotten."*



TECHNIQUES USED IN NLP

thank you
hvala
shukran
kiitos
tack
mh'gōi
tānan
dankie
ačiū
grazzi
hvala
danku
danke
obrigado
tak
gràcies
grazie
merci
takkhvala
arigatô
takk
xièxiè
gracias
terima kasih
nandri
choukrane
mahalo
děkuji
faleminderit
благодаря
благодарам
danke terima kasih
tak
merci
hvala