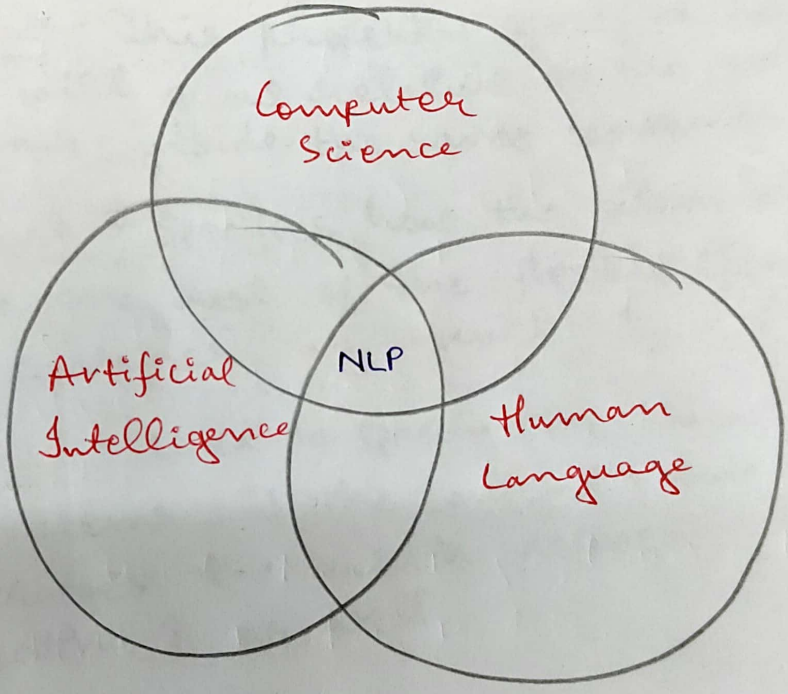
NATURAL LANGUAGE PROCESSING

LECTURE-1

↳ NLP is a subfield of linguistics, computer science and ~~human~~ ~~language~~ artificial intelligence concerned with the interactions b/w computers and human language, in particular way to program computers to process and analyze large amounts of natural language data.



Need for NLP

↳ In neuropsychology, linguistics, and the philosophy of language, a natural language or ordinary language is any language that has evolved naturally in humans through use and repetition without conscious planning or premeditation. Natural language can take different forms, such as speech or signing.

↳ They are distinguished from constructed and formal languages such as those used to program computers or to study logic.

Real World Applications

(1) Contextual Advertisements

(2) Email Clients - spam filtering, smart reply

(3) Social Media - removing adult content, opinion mining

(4) Search Engines

(5) Chatbots.

Common NLP Tasks

↳ Text/Document Analysis
↳ Sentiment Analysis
↳ Information Retrieval
↳ Parts of speech Tagging
↳ Language Detection and Machine Translation
↳ Conversational Agents
↳ Knowledge Graph and QA System
↳ Text Summarization.
↳ Topic modelling
↳ Text Generation
↳ Spell Checking and Grammar Correction
↳ Text Parsing.
↳ Speech to Text

Approaches to NLP

(1) Heuristic Technique (1950-1990)

↳ It is an approach to problem solving or self discovery that employs a practical method that is not guaranteed to be optimal, perfect, or rational, but is nevertheless sufficient for reaching an immediate, short-term goal or approximation.

↳ Examples - Regular Expressions, WordNet, Open Mind Common Sense.

(2) Machine Learning Approach.

↳ It is an area of Artificial Intelligence, where machines learn (predict) tasks based on previous experience (data). Basically, machine learning algorithms are developed using two types of datasets. Training dataset is used to train the model.

↳ Algorithms used - Naive Bayes, Logistic Regression, SVM, LDA, Hidden Markov Models.

(3) Deep Learning Approach

↳ Retains Sequential Order.
↳ Feature Generation is automatic
} Advantages in DL Models.

↳ Architectures used - RNN, LSTM, GRU/CNN, Transformers, Autoencoders

Challenges in NLP

(1) Ambiguity

→ I saw the boy on the beach with my binoculars.

→ I have never tasted a cake quite like that one before!

(2) Contextual words

→ I ran to the store because we ran out of milk.

(3) Colloquialisms and slang

→ Piece of cake

→ Pulling your leg.

(4) Synonyms

(5) Irony, Sarcasm and Tonal Difference

↳ That's just what I needed today!

(6) Spelling errors

(7) Diversity

(8) Creativity

↳ Poems, dialogue, scripts.