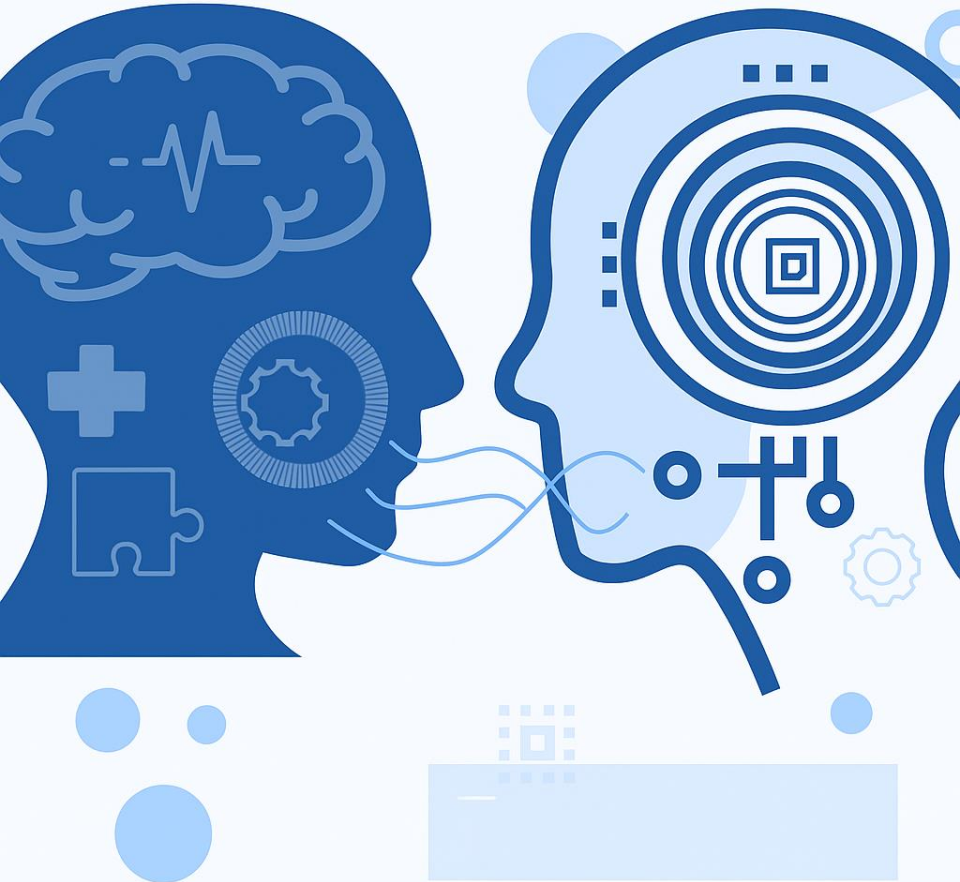


NLP

Natural
Language
Processing



NATURAL LANGUAGE PROCESSING (NLP)

PMDS606L

MODULE 1

LECTURE 1

Dr. Kamanasish Bhattacharjee

Assistant Professor

Dept. of Analytics, SCOPE, VIT



A

Apple



B

Ball



C

Cat



D

Dog



TYPES OF DATA

- Tabular Data
- Image Data
- Signals
- Text

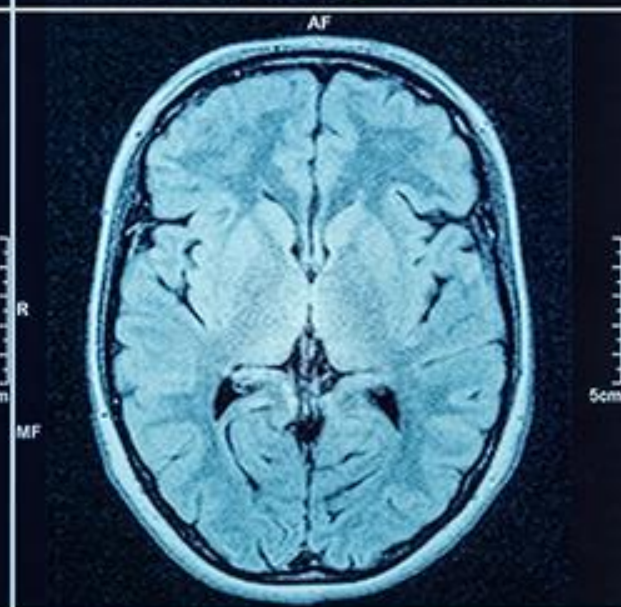
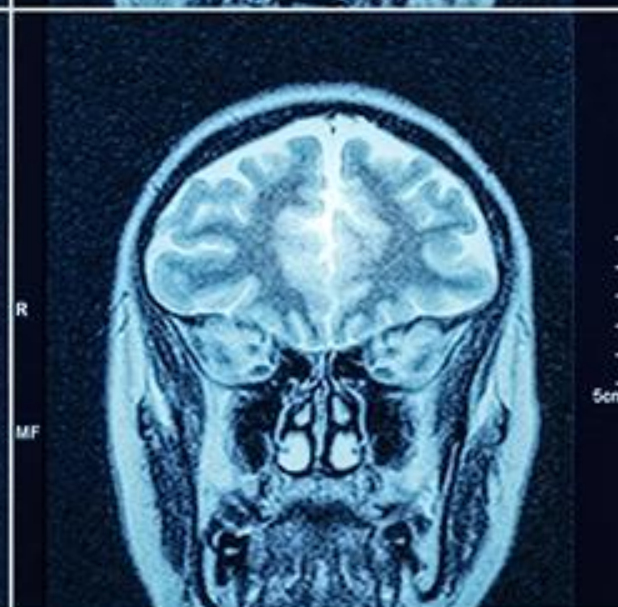
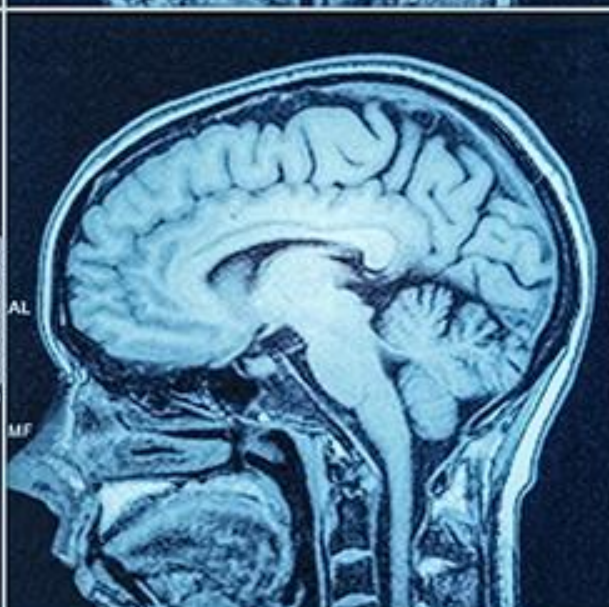
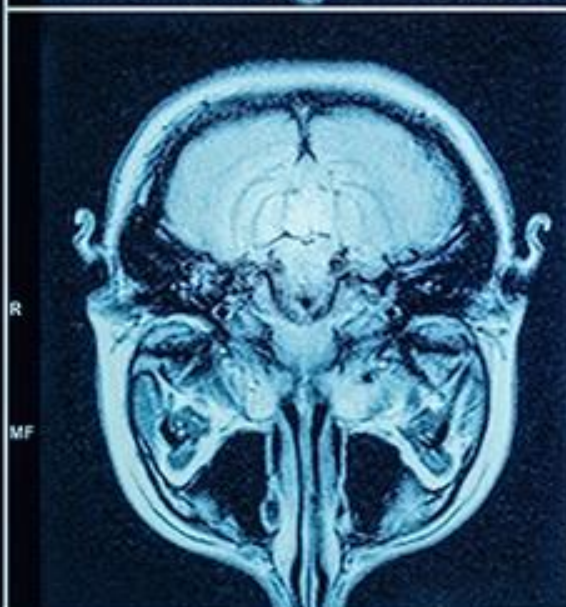
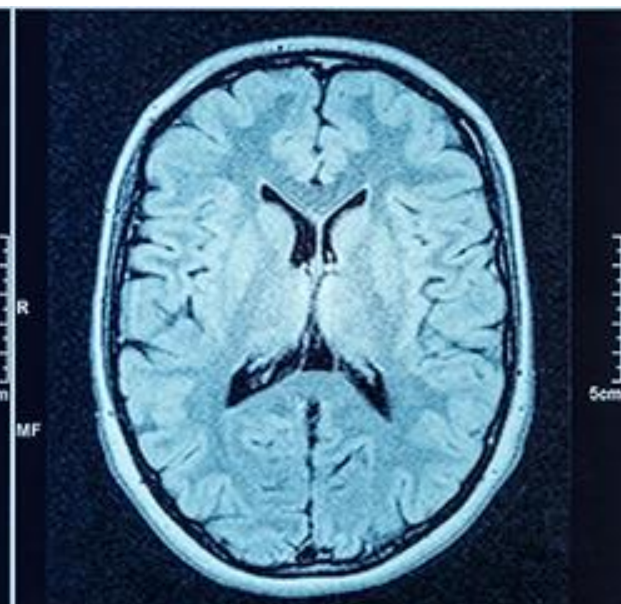
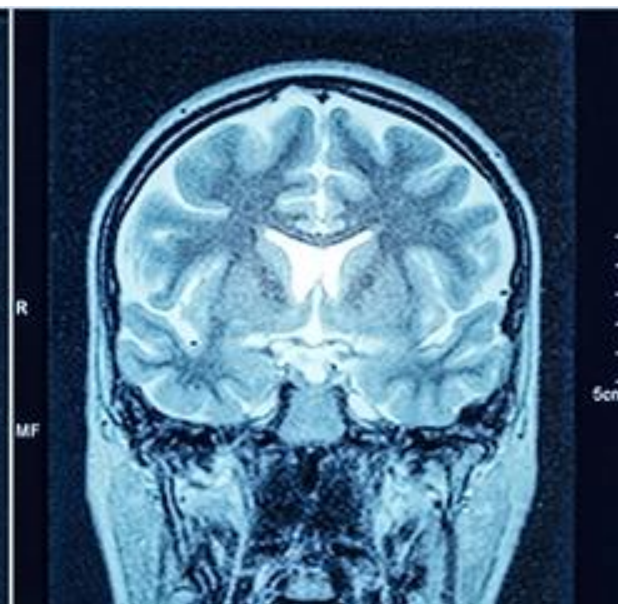
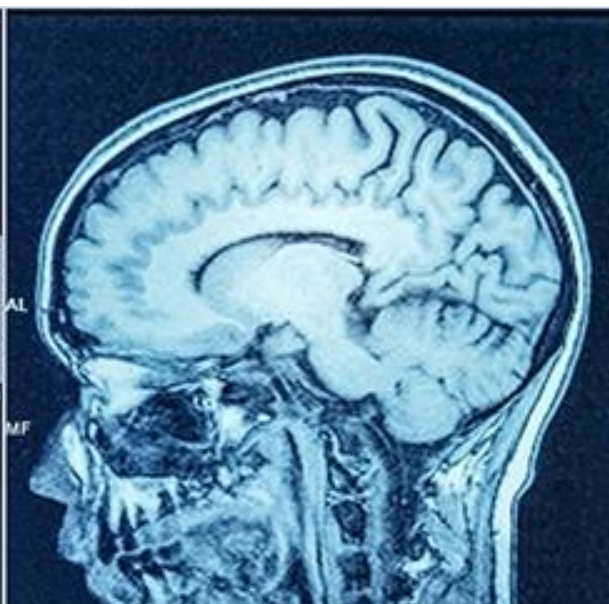
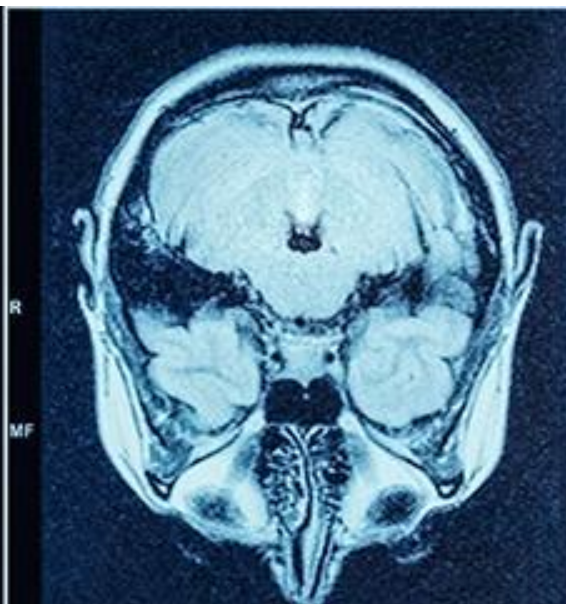
Tabular Data

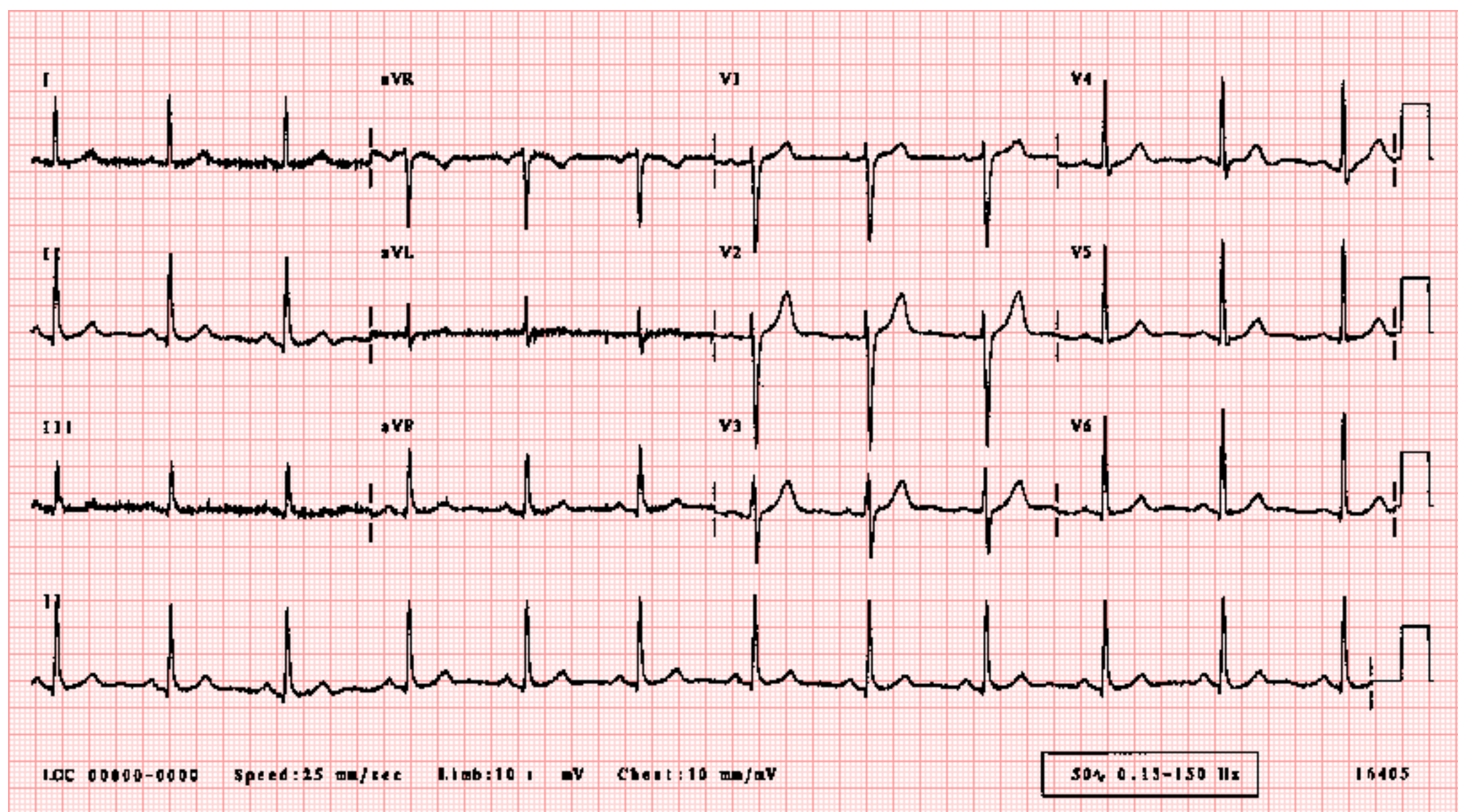
columns = attributes for those observations

Rows = observations



Player	Minutes	Points	Rebounds	Assists
A	41	20	6	5
B	30	29	7	6
C	22	7	7	2
D	26	3	3	9
E	20	19	8	0
F	9	6	14	14
G	14	22	8	3
I	22	36	0	9
J	34	8	1	3





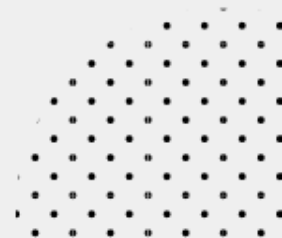


TEXTUAL DATA

ADVANTAGES

- Textual data is easily accessible, as you can find it in books, articles, web pages, and other sources.
- Collecting and storing textual data is relatively inexpensive compared to images and videos.
- Textual data can provide a wealth of information about a particular topic or area of research.
- Analyzing textual data can help uncover patterns, trends, and correlations that are not immediately visible.
- Textual data can be used to gain insights into the sentimental data of specific groups (like users who contact customer support).

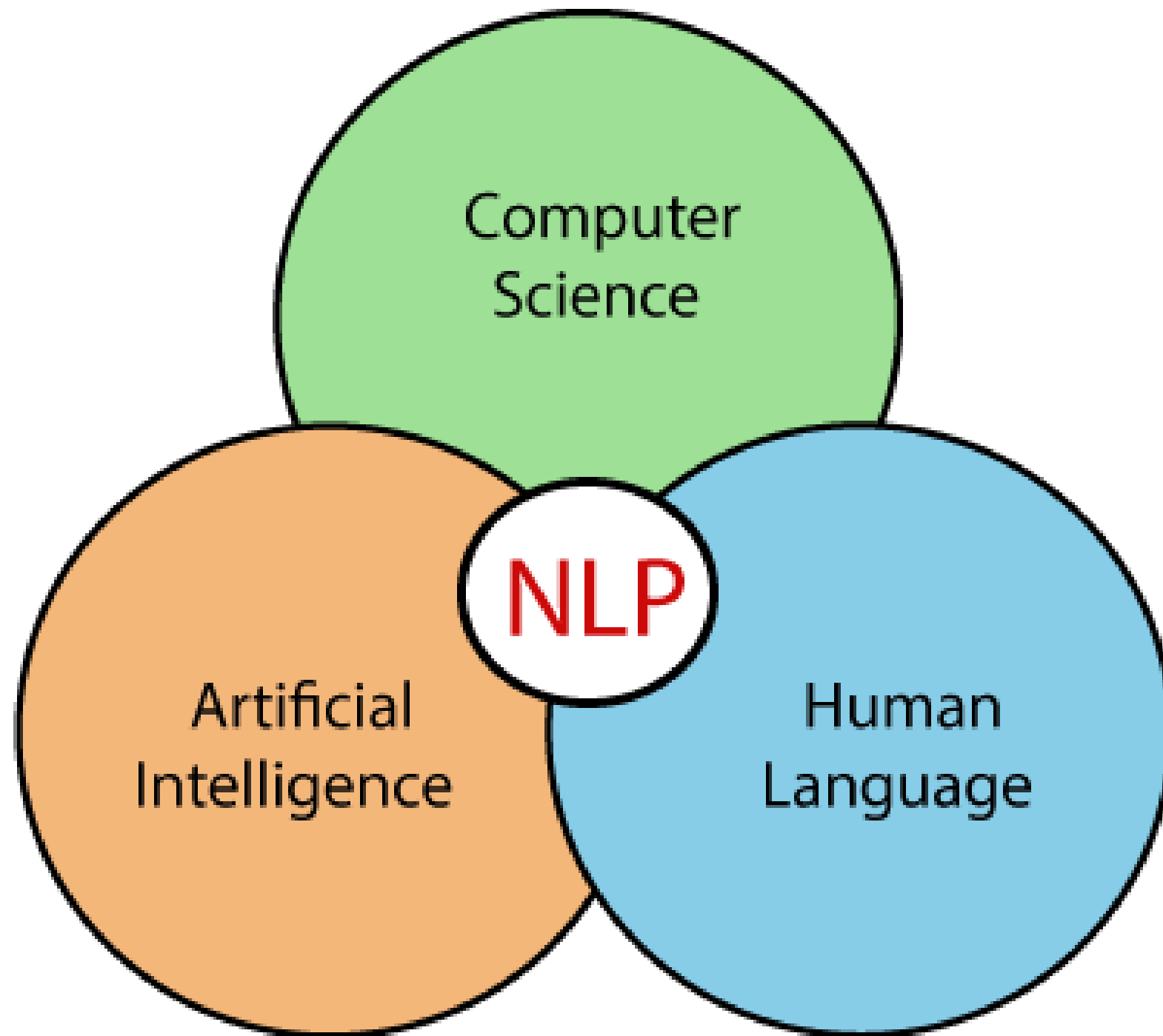
DISADVANTAGES

- Textual data is often unstructured and easier to analyze with specialized tools such as natural language processing (NLP).
 - Textual data is susceptible to bias due to the subjective nature of human language.
 - It's time-consuming to extract meaningful insights from extensive collections of textual data manually.
 - Textual data may contain errors or inaccuracies, leading to inaccurate or misleading conclusions if not detected and corrected.
- 



NLP TASKS

- **Understand** human language (text or speech).
- **Translate** between languages (e.g., English to French).
- **Generate** human-like language (e.g., writing summaries or stories).
- **Extract** information and meaning from large volumes of unstructured text.
- **Interact** via speech or text (e.g., chatbots, voice assistants).



English ► Hindi

Waheeda Rehman on the set of Bees Saal Baad (1962).

वहीदा रहमान मधुमक्खियों साल बाद (1962) के सेट पर।

⚙ Rate this translation



The *cat* drank the milk because **it** was hungry.

The cat drank the *milk* because **it** was sweet.

Sapir–Whorf hypothesis / Whorf hypothesis / Whorfianism: the structure of a language influences its speakers' worldview or cognition, and thus individuals' languages determine or shape their perceptions of the world



- In the **Russian language**, there are separate words for **light blue** ("голубой" – *goluboy*) and **dark blue** ("синий" – *siniy*).
 - Russian speakers were faster at distinguishing light and dark shades of blue than English speakers, who use just the word "blue" for both.
- The **Kuuk Thaayorre people** (an Aboriginal group in Australia) don't use "left" or "right"—they use cardinal directions (north, south, east, west) for everything. Instead of saying, "The cup is to your left," they say, "The cup is to your southwest."
 - This makes them extremely good at always knowing **which direction they're facing**, because their **language requires constant spatial awareness**.

GEOFFREY
RUSH THE COLIN
FIRTH
KING'S
SPEECH
HELENA
BONHAM CARTER

