

# Natural Language Processing (NLP)

---

Lecture 21 – HCCDA-AI

# Overview

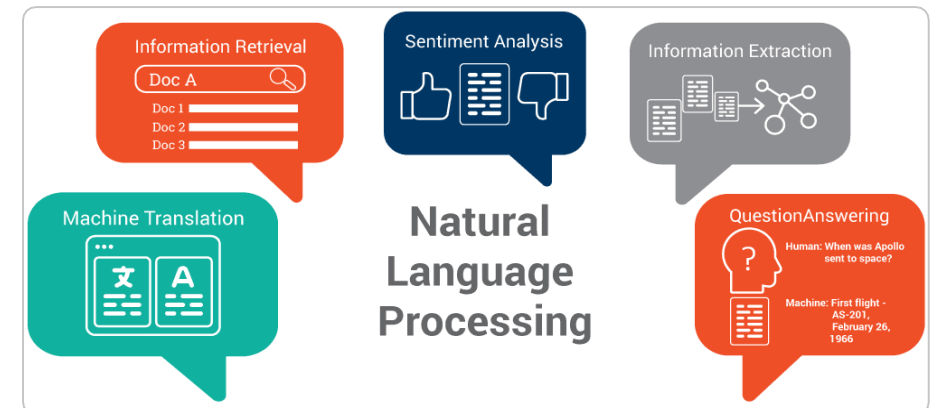
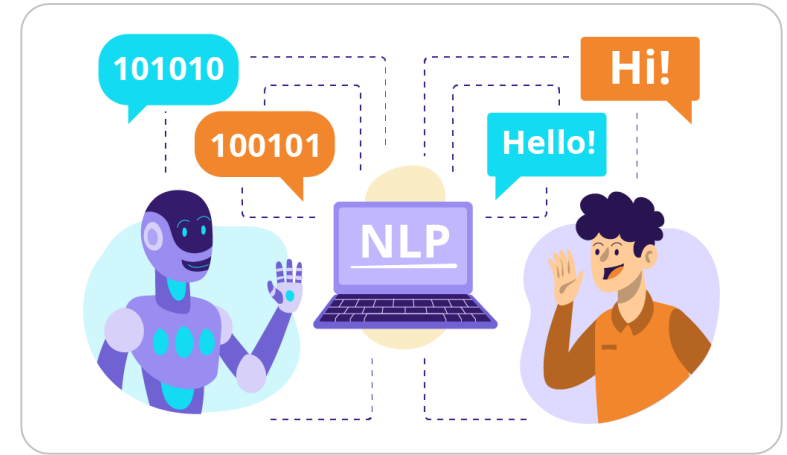
- **Natural Language Processing (NLP)**
- History and Evolution of NLP
- Milestones in Modern Neural NLP
  - Neural Machine Translation
  - Free-Text Question Answering
  - GPT-2 (2019)
  - ChatGPT, GPT-4 and Beyond
- **Different Approaches of NLP**
  - Rule-Based NLP (1950s – 1980s)
  - Statistical NLP (1980s – 1990s)
  - Machine Learning in NLP (1990s – 2010s)
  - Deep Learning in NLP (2013 – Present)
- Core NLP Tasks (with Examples)
- Sentiment Analysis
- **NLP Pipeline:** From Raw Text to Insight
  - Text Pre-processing
  - Feature Extraction Techniques
  - Modeling/Learning
- **Lab:** Sentiment Analysis with Word2Vec + LSTM (PyTorch)

# Introduction to NLP

- Language helps us talk to each other, and now to computers too!
- Natural Language Processing (NLP) teaches machines to understand, interpret, and generate human language.
- Field at the intersection of linguistics and machine learning focused on understanding human language.
- Aims to understand context, not just words.

## Why NLP Matters:

- Makes **chatbots** like ChatGPT and Siri possible
- Powers **search engines** like Google
- Helps **translate languages** (Google Translate)
- Finds out what **people feel in reviews** (sentiment analysis)
- NLP is foundational to advanced AI systems

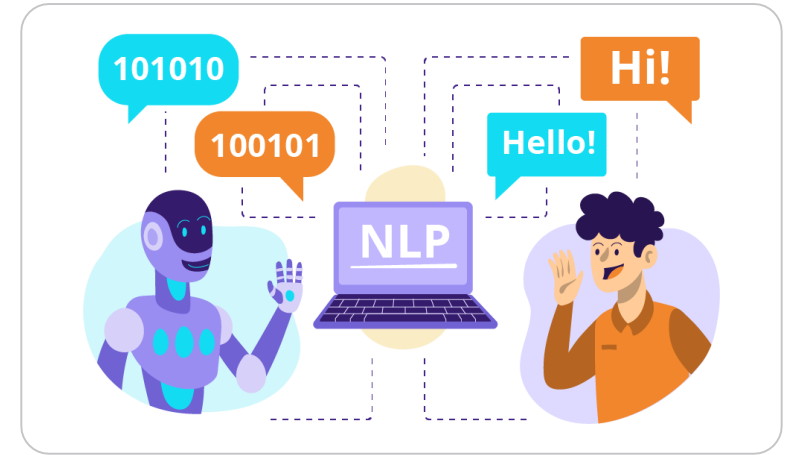


# What is NLP

- Study of computational approaches to processing natural languages.
  - **Processing includes:**
    - Acquiring language data
    - Representing information
    - Storing text and speech
    - Understanding meaning
    - Characterizing language patterns
    - Generating new language
- Natural languages refer to human languages.

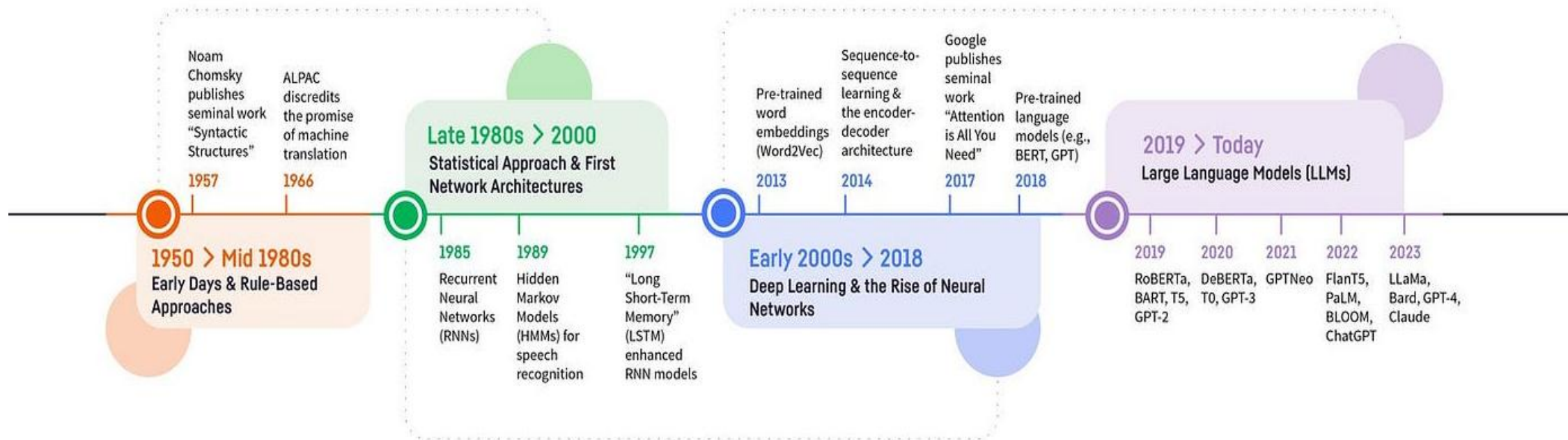
## Goal of NLP:

- Enable machines to understand and generate human language.
- Facilitate human-computer interaction through natural language.
- Develop systems that can process and analyze large amounts of text data.



# History and Evolution of NLP

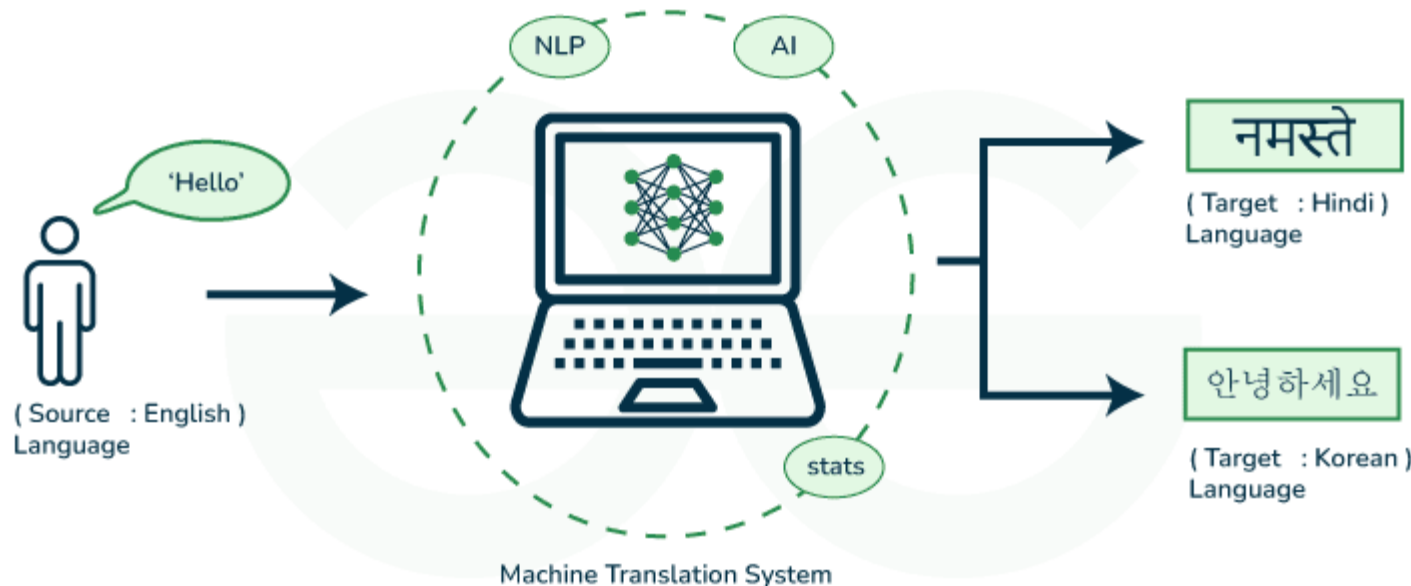
Era	Approach	Example	Limitation
1950s – 1980s	Rule-Based	ELIZA chatbot	Brittle, hard to scale
1980s – 1990s	Statistical NLP	HMMs for POS tagging	Requires lots of data
1990s – 2010s	Machine Learning	SVMs for sentiment analysis	Feature engineering required
2010s – Now	Deep Learning	RNNs, LSTMs, Transformers	Requires compute, data



# Milestones in Modern Neural NLP

## 1. Neural Machine Translation: The First Major Success of Neural NLP.

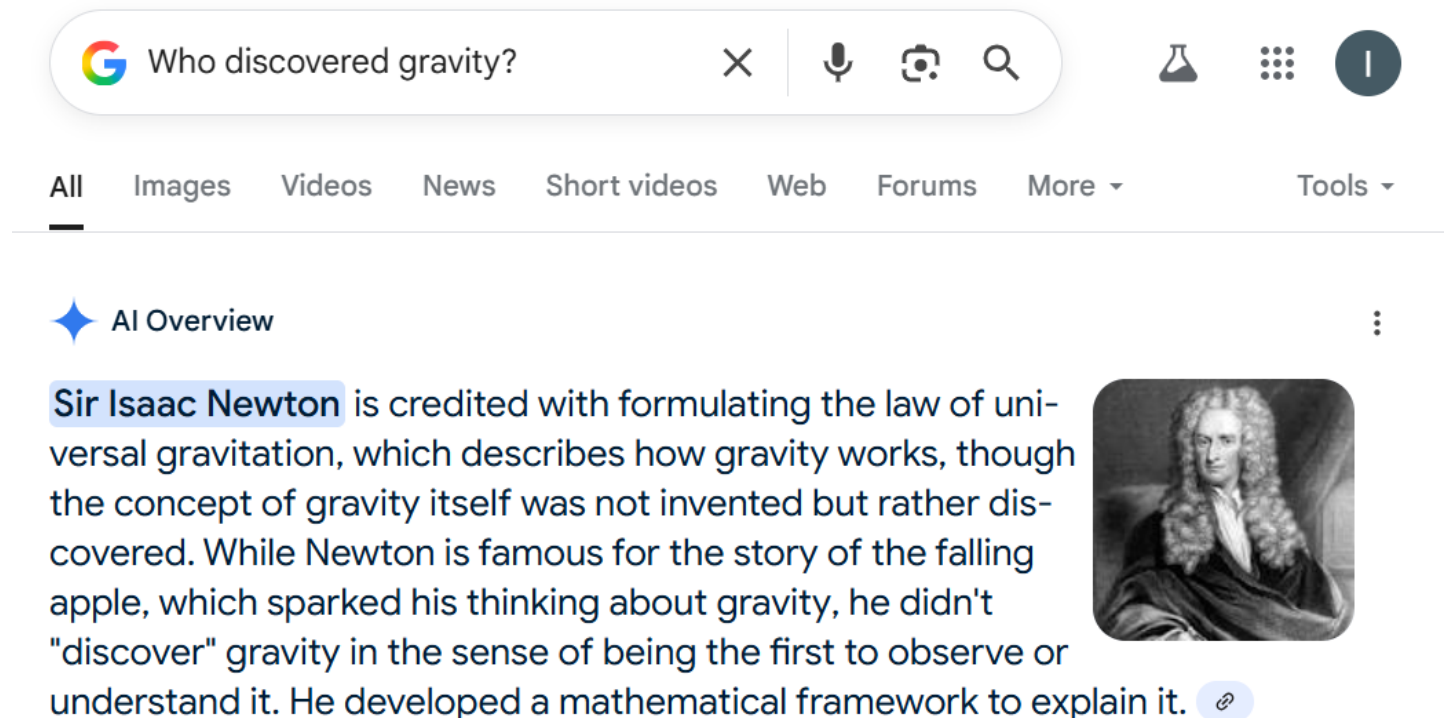
- **2014:** Seq2Seq (RNN) by Google Brain
- **2015:** Attention mechanism introduced
- **2016:** Google Translate adopts NMT



# Milestones in Modern Neural NLP

## 2. Free-Text Question Answering: The Next Generation of Search.

- Move from keyword search → direct answer generation
- Powered by neural models (e.g., BERT, early QA systems).
- Understands context, not just terms.



The screenshot shows a Google search interface. The search bar contains the text "Who discovered gravity?". Below the search bar, there are tabs for "All", "Images", "Videos", "News", "Short videos", "Web", "Forums", and "More". The "All" tab is selected. Below the tabs, there is a section titled "AI Overview" with a blue star icon. The text in this section reads: "Sir Isaac Newton is credited with formulating the law of universal gravitation, which describes how gravity works, though the concept of gravity itself was not invented but rather discovered. While Newton is famous for the story of the falling apple, which sparked his thinking about gravity, he didn't 'discover' gravity in the sense of being the first to observe or understand it. He developed a mathematical framework to explain it." To the right of the text is a portrait of Sir Isaac Newton. At the bottom right of the text is a small blue icon with a link symbol.



# Milestones in Modern Neural NLP

## 3. GPT-2 (2019): A Step Toward Foundation Models

### GPT-2 (2019): An early step towards foundation models OpenAI

PROMPT  
(HUMAN-WRITTEN)

*A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.*

MODEL COMPLETION

The incident occurred on the downtown train line, which runs from Covington and Ashland stations.

In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.

“The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,” said Tom Hicks, the U.S. Energy Secretary, in a statement. “Our top priority is to secure the theft and ensure it doesn’t happen again.”

The stolen material was taken from the University of Cincinnati’s Research Triangle Park nuclear research site, according to a news release from Department officials.

Stanford




# Milestones in Modern Neural NLP

## 4. ChatGPT, GPT-4, and Beyond: Rise of General-Purpose Language Models

### ChatGPT, GPT-4, and more



**A** Hey please draft a polite mail to explain my boss Jeremy that I would not be able to come to office for next 2 days because my 9 year son Peter is angry with me that I am not giving him much time. So I have plans to take him out for a trip. I will return and complete my work by meeting deadline.

 Dear Jeremy,

I hope this email finds you well. I'm writing to let you know that I will not be able to come into the office for the next two days.

My 9-year-old son Peter is feeling upset with me because he feels like I haven't been giving him enough attention lately. In order to make it up to him, I've decided to take him on a little trip to help us reconnect.

I understand that this may cause some inconvenience, but I promise to make up for any lost time by working extra hard to meet any deadlines when I return.

Thank you for your understanding.

Best regards,  
[Your Name]

# Approaches to NLP

## 1) Rule-Based NLP (1950s–1980s)

- In the early days of NLP, systems were primarily rule-based. Linguists and programmers manually created a set of rules to process language.
- **Example: ELIZA** chatbot using pattern matching.
- Used in early grammar checkers, simple chatbots.

## 2. Statistical NLP (1980s–1990s)

- The 1980s and 1990s saw a shift from rule-based systems to statistical methods. This approach utilized large amounts of data and statistical techniques to understand and generate language.
- **Example:** Hidden Markov Models (HMMs) for **POS tagging**.
- Used in early speech recognition, translation.

```
Welcome to
      EEEEE LL   IIII ZZZZZ AAAAA
      EE   LL   II   ZZ   AA   AA
      EEEEE LL   II   ZZZ  AAAAAA
      EE   LL   II   ZZ   AA   AA
      EEEEE LLLLL IIII ZZZZZ AA   AA

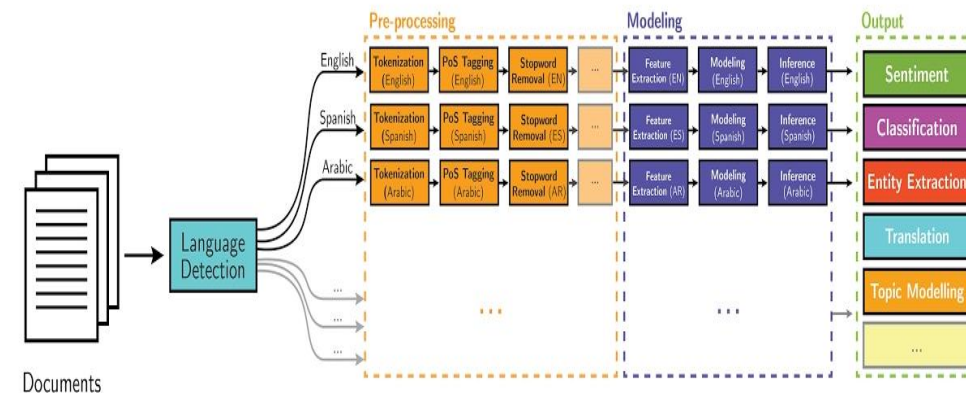
Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

# Approaches to NLP

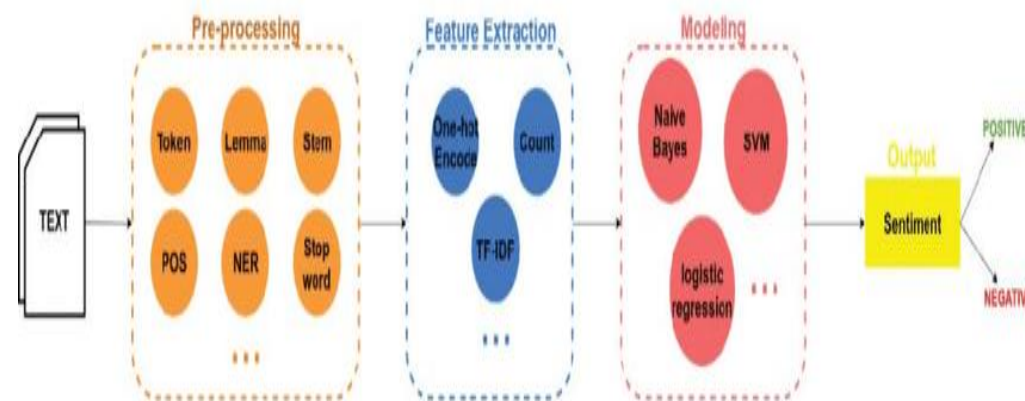
## 3. Machine Learning in NLP (1990s–2010s)

- The late 1990s and early 2000s marked the introduction of machine learning in NLP. Models learn patterns from labeled data.
- **Techniques:** Naive Bayes, SVMs, Decision Trees.
- **Requires:**
  - Feature engineering
  - Clean, labeled data



## 4. Deep Learning in NLP (2013–Present)

- **Key Models:**
  - **RNNs/LSTMs:** Good at handling sequences.
  - **CNNs:** Capture local patterns in text.
  - **Transformers:** Now dominant in modern NLP.
- **The Rise of Large Language Models (LLMs):**
  - Revolutionized NLP with models like GPT and Llama.
  - Trained on massive text data to understand and generate human-like text.
  - Shifted from task-specific models to general-purpose models.
- **Applications:**
  - Machine Translation
  - Text Generation
  - Q&A Systems
  - Summarization



# Core NLP Tasks (with Examples)

Task	What it does	Example
<b>Tokenization</b>	Split text into words	“Pakistan Zindabad” → [“Pakistan”, “Zindabad”]
<b>Text Classification</b>	Assigning categories to text	Spam detection in emails
<b>Sentiment Analysis</b>	Detecting mood	“This movie was amazing!” → Positive
<b>Machine Translation</b>	Language to language	English → Urdu
<b>Named Entity Recognition</b>	Find names, places, etc.	Imran Khan, 2024, Islamabad
<b>Text Summarization</b>	Generating concise summaries	Abstract generation from research articles

# Sentiment Analysis

- **Definition:**

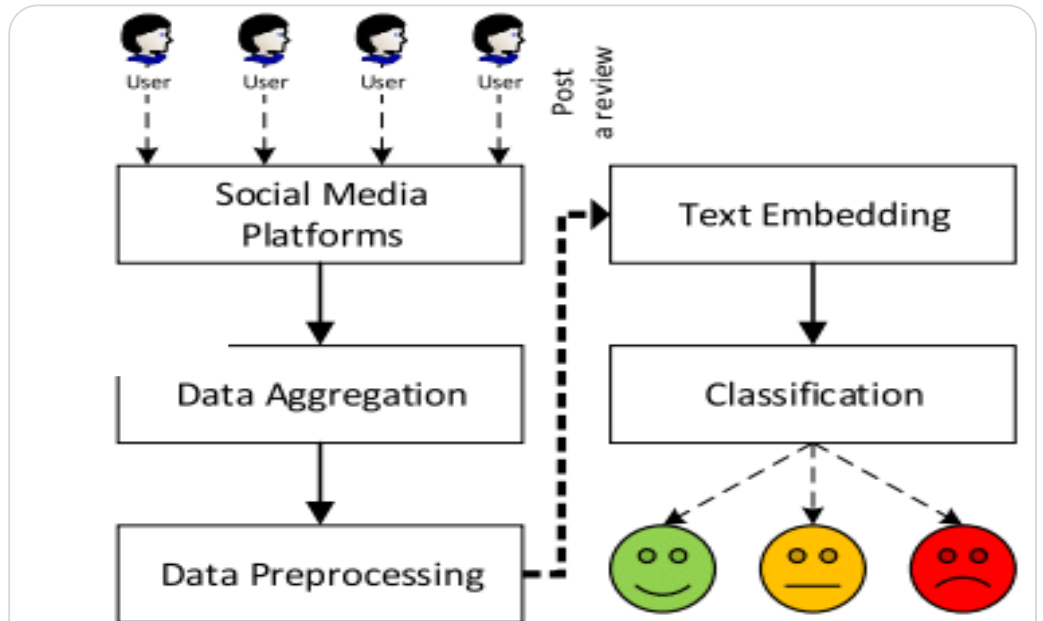
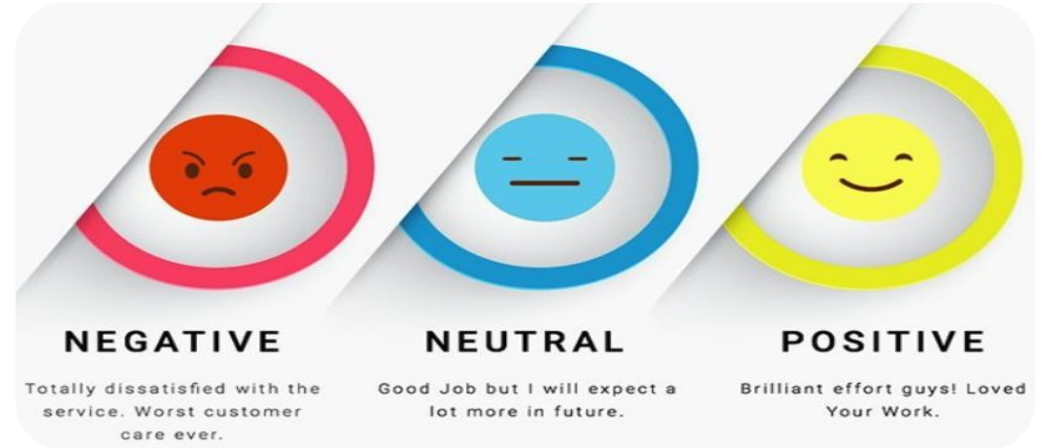
- Using NLP to find out if text expresses a positive, negative, or neutral opinion about a topic, product, or service.

- **Example:**

- **Sentence:** “I love the new design of this app, but I hate how it crashes frequently.”
- **Positive Sentiment:** “I love the new design of this app.”
- **Negative Sentiment:** “I hate how it crashes frequently.”

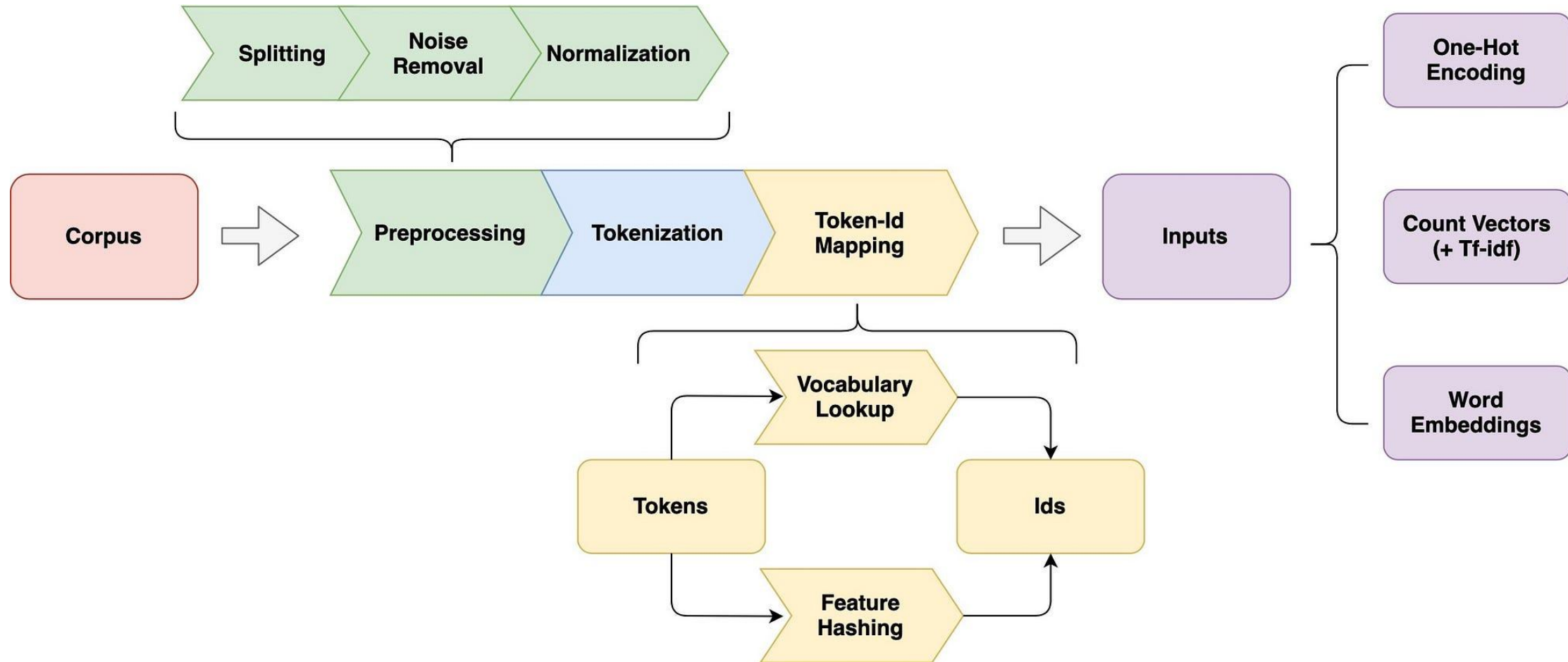
- **Use Case:**

- **Brand Management:** Companies track tweets, reviews, and posts to see how people feel about a new product. This helps them fix issues quickly and keep customers happy.



**FIGURE 1.** Basic steps of sentiment analysis on social media.

# NLP Pipeline: From Raw Text to Insight



# NLP Pipeline: From Raw Text to Insight

## 1. Text Preprocessing

- Essential to prepare raw language for modeling.

Step	Purpose
Tokenization	Split text into words/subwords (tokens)
Removing stop words	Remove common, non-informative words (e.g., “is”, “the”, “and”)
Stemming/Lemmatization	Reduce words to their root or base form. Example: “running” → “run”
Lowercase everything	“NLP” → “nlp”
Punctuation Removal	Clean text
POS Tagging	Assigning parts of speech (noun, etc ) to each token.
Named Entity Recognition	Extract named entities
Spell Checking	Correct spelling errors
Handling Duplicates/Missing	Ensure clean and complete input

### Example:

- Sentence: “*It was the best of times.*”
- Tokens: [‘it’, ‘was’, ‘the’, ‘best’, ‘of’, ‘times’]

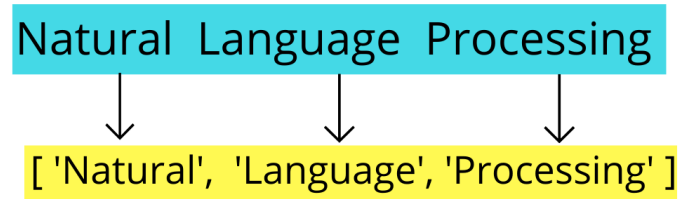


# NLP Pipeline: From Raw Text to Insight

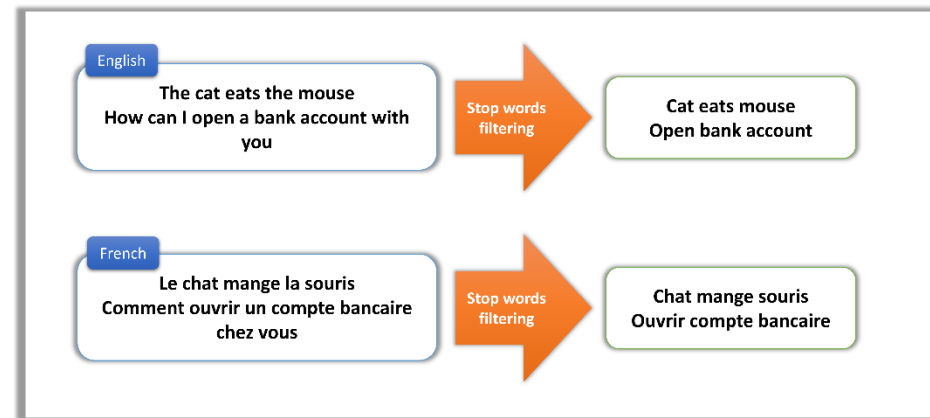
## 1. Text Preprocessing

- Essential to prepare raw language for modeling.

### Tokenization



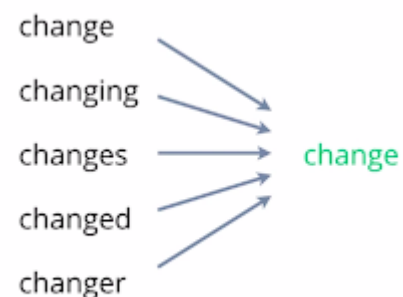
### Removing Stop words



### Stemming



### Lemmatization



### Text Normalization

	Rating	Reviews	Clean_Reviews
0	5	I feel so LUCKY to have found this used (phone...	i feel so lucky to have found this used phone ...
1	4	nice phone, nice up grade from my pantach revu...	nice phone nice up grade from my pantach revue...
2	5	Very pleased	very pleased
3	4	It works good but it goes slow sometimes but i...	it works good but it goes slow sometimes but i...
4	4	Great phone to replace my lost phone. The only...	great phone to replace my lost phone the only ...

# NLP Pipeline: From Raw Text to Insight

## 2. Feature Extraction Techniques / Vectorization:

- Transform words or documents into numerical representations for machine learning models.

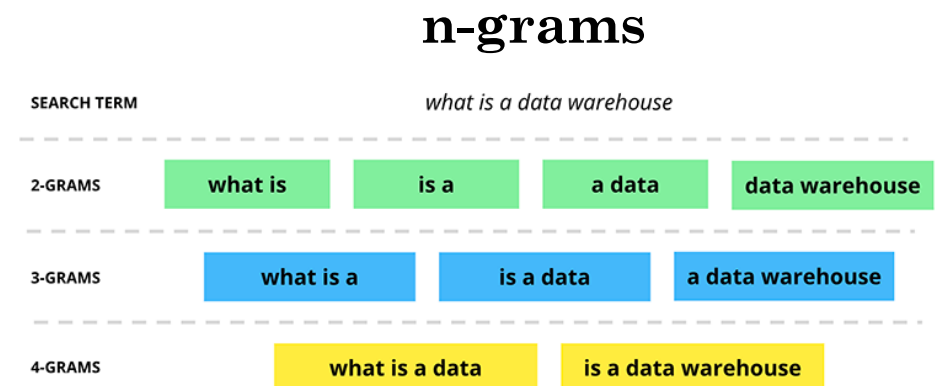
Technique	Description
Bag-of-Words	Counts word occurrences in a document.
TF-IDF	Weighs words by importance across documents.
n-Grams	Groups of 2 or 3 words together to keep some word order and context.
Word2Vec / GloVe	Learns dense vector representations capturing word meaning and context.

### Example:

**Sentence:** *“it was the best of times”*

**Vocabulary:** [it, was, the, best, of, times]

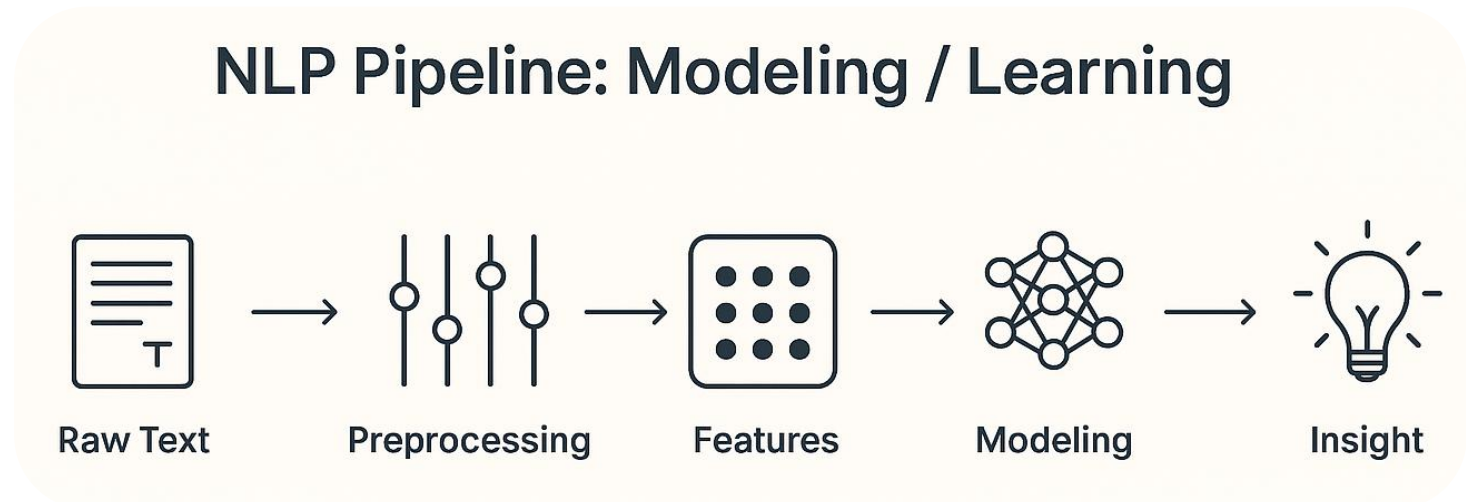
**BoW Vector:** [1, 1, 1, 1, 1, 1]



# NLP Pipeline: From Raw Text to Insight

## 3. Modeling/Learning:

- Using machine learning or deep learning algorithms to learn patterns from extracted text features.
- **Goal:** Predict, classify, or generate meaningful language output.
- **Common Approaches:**
  - **Traditional ML:** Logistic Regression, Naive Bayes, SVM (fast, simple, works well with BoW/TF-IDF).
  - **Deep Learning:** RNNs, LSTMs, GRUs (capture sequence).
  - **Transformers:** BERT, GPT (state-of-the-art for NLP tasks).



Thank You