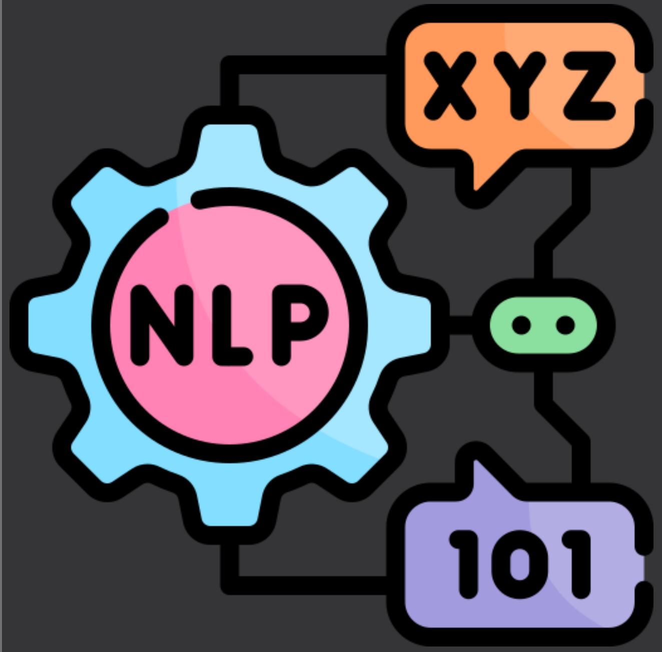


Natural Language Processing (NLP)

Lecture 21 – HCCDA-AI

Imran Nawar

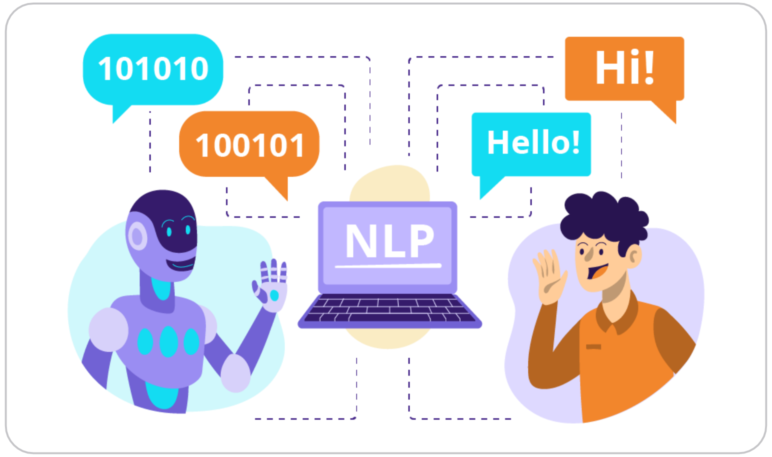
23 August 2025

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)

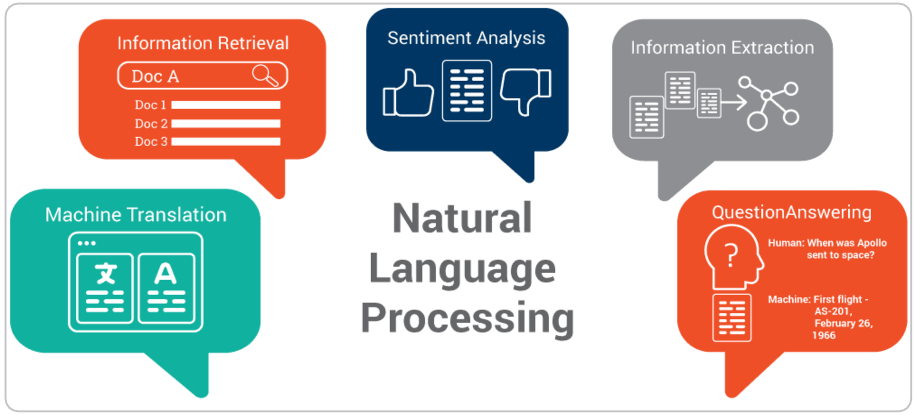
Introduction to NLP

- 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)

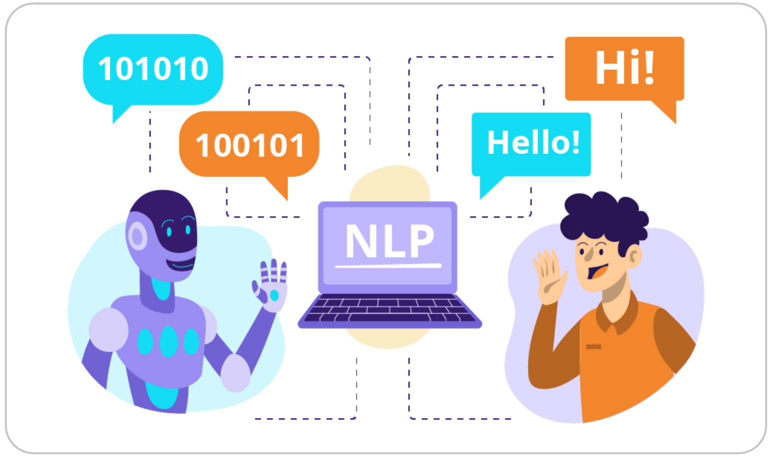


- NLP is foundational to advanced AI systems

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.

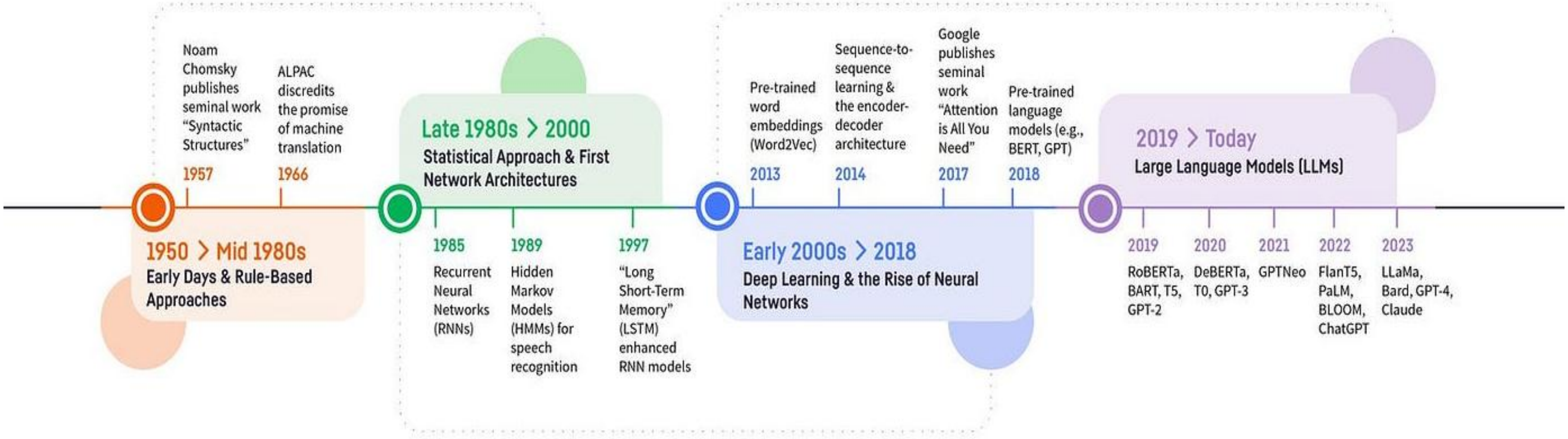


Goal of NLP:

- 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.

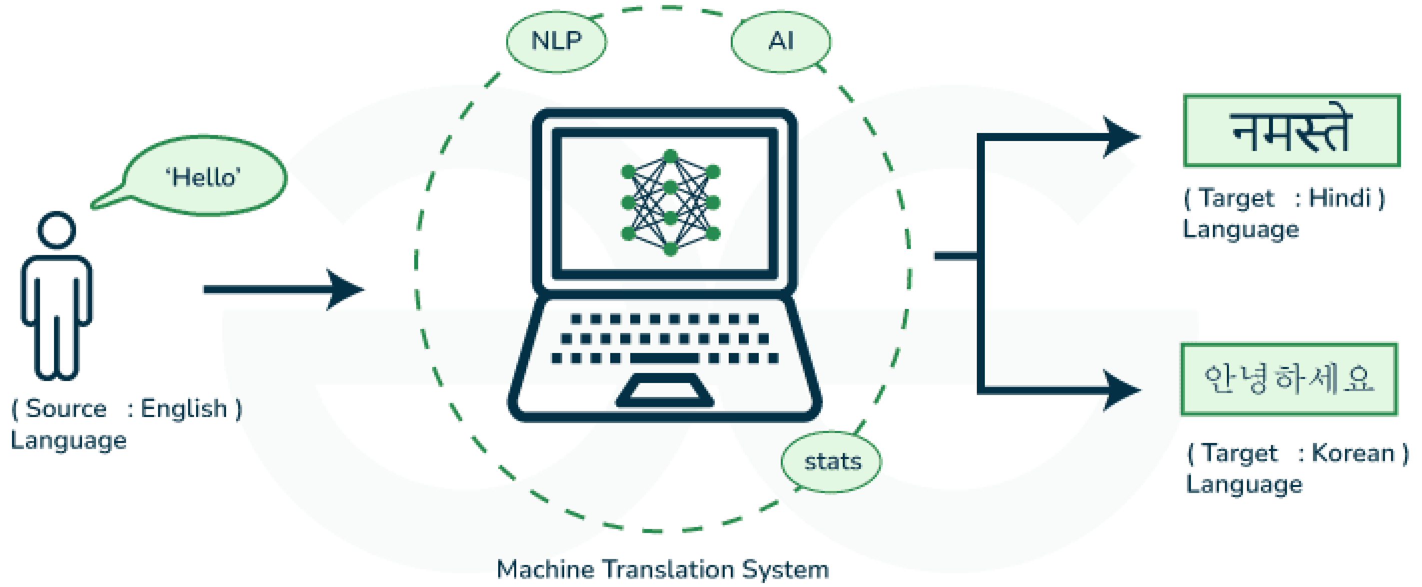
History and Evolution of NLP

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



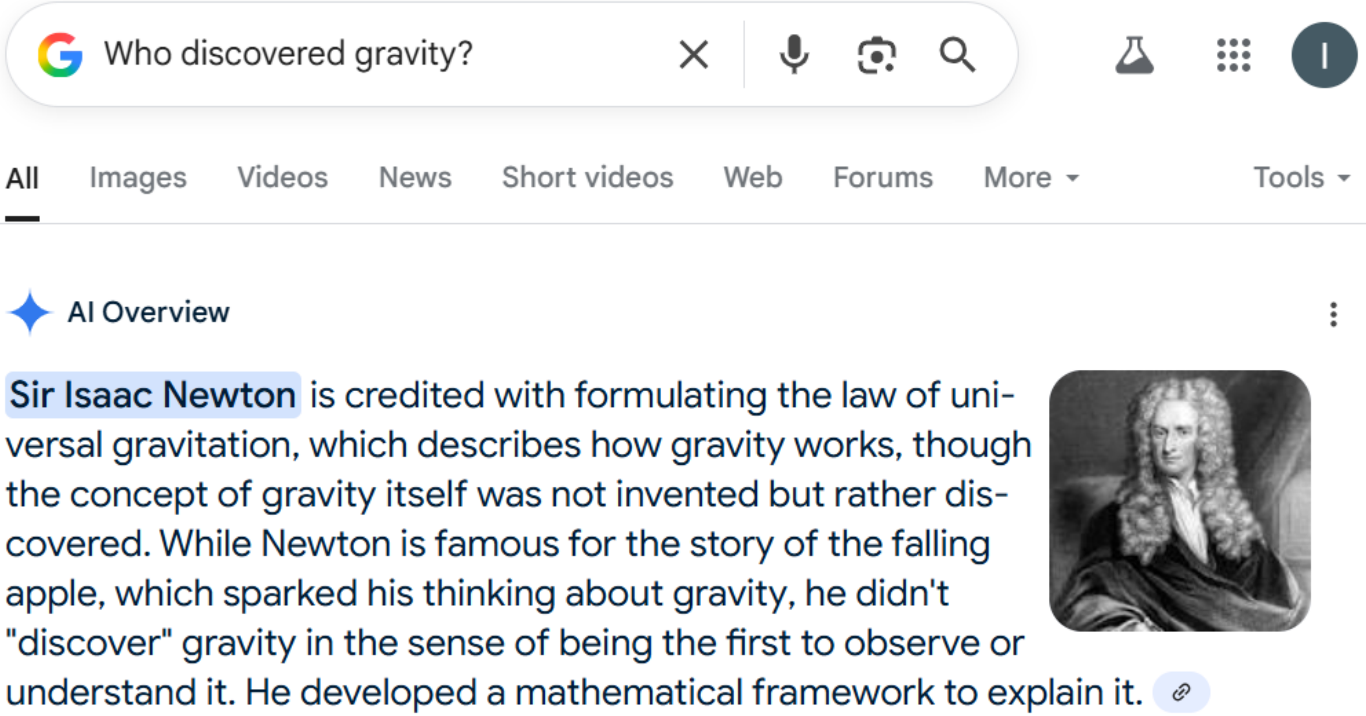
Milestones in Modern Neural NLP

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



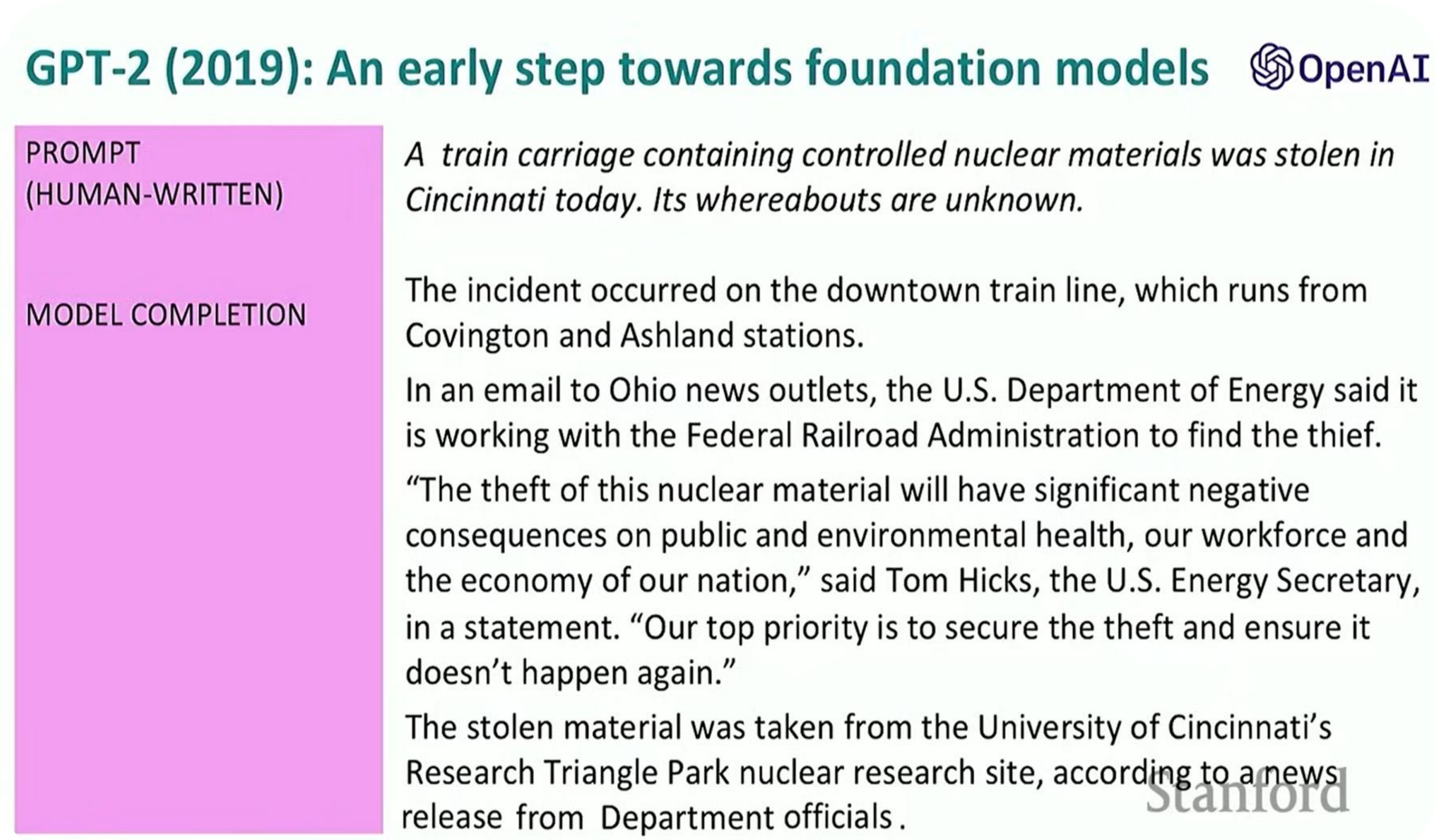
Milestones in Modern Neural NLP

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.



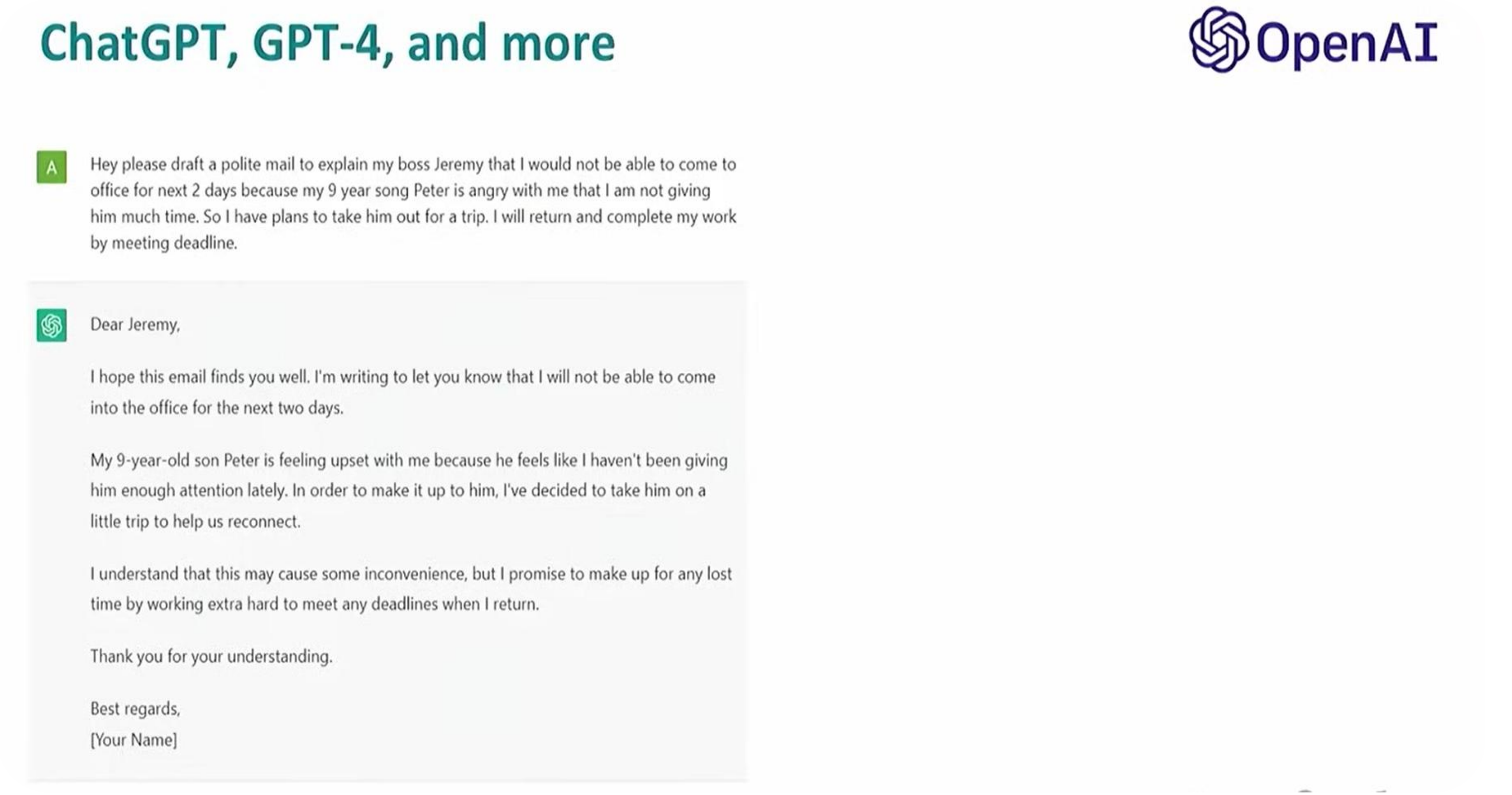
Milestones in Modern Neural NLP

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



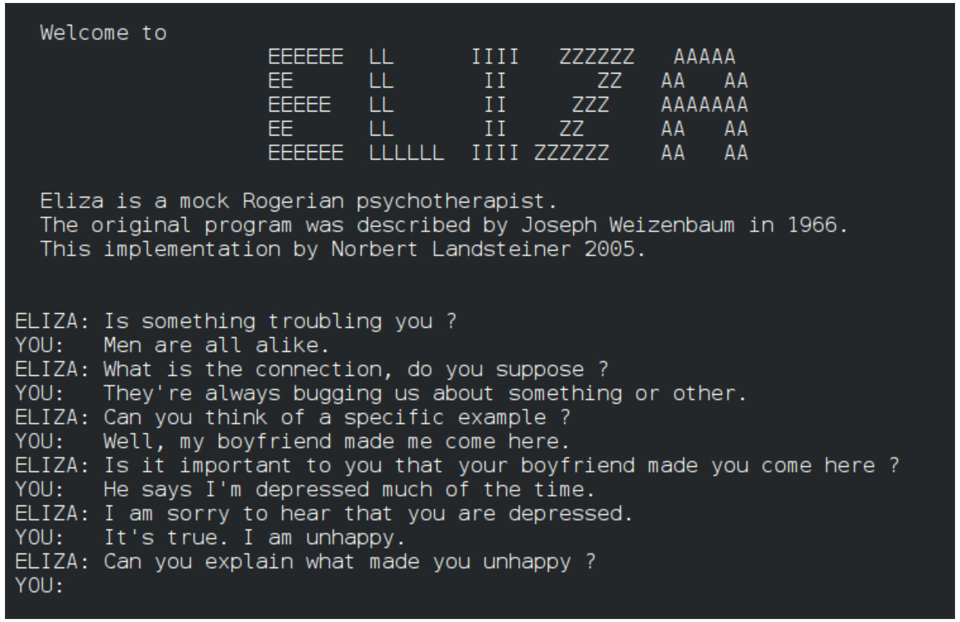
Milestones in Modern Neural NLP

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



Approaches to NLP

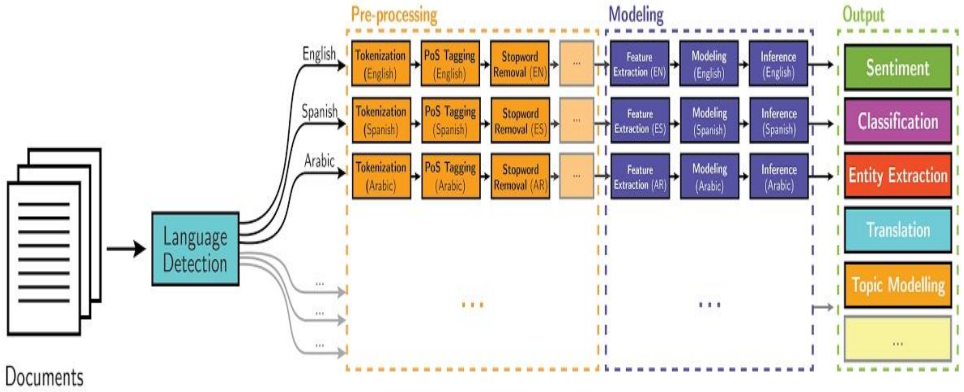
1) **Rule-Based NLP (1950s–1980s)**
   - In the early days of NLP, systems were primarily rule-based. Linguists and programmers manually created a set of rules to process language.
   - **Example: ELIZA** chatbot using pattern matching.
   - Used in early grammar checkers, simple chatbots.



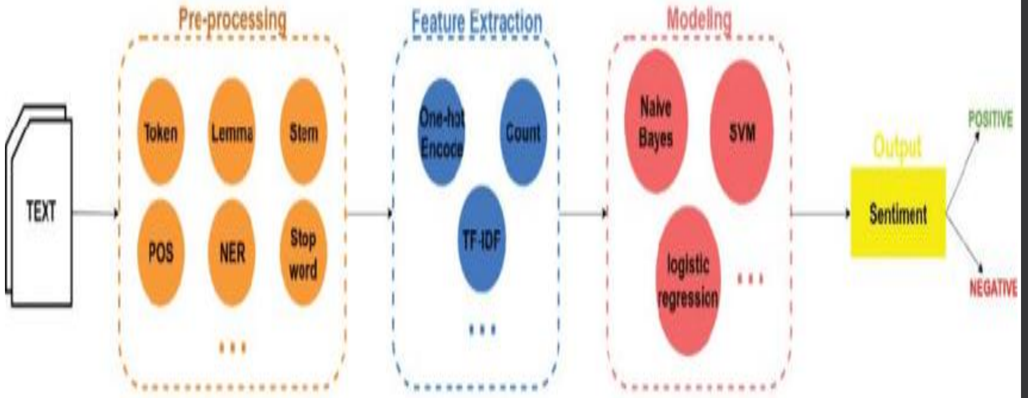
2. **Statistical NLP (1980s–1990s)**
   - The 1980s and 1990s saw a shift from rule-based systems to statistical methods. This approach utilized large amounts of data and statistical techniques to understand and generate language.
   - **Example:** Hidden Markov Models (HMMs) for **POS tagging**.
  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

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

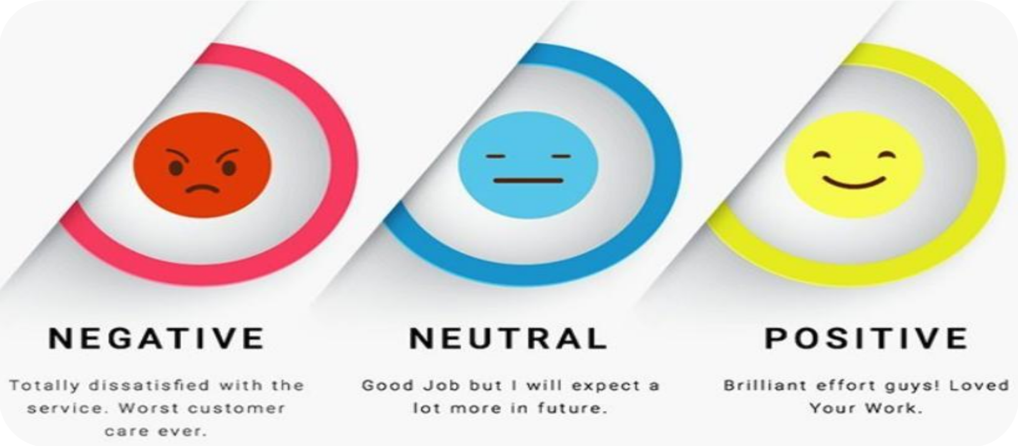
Sentiment Analysis

- **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.



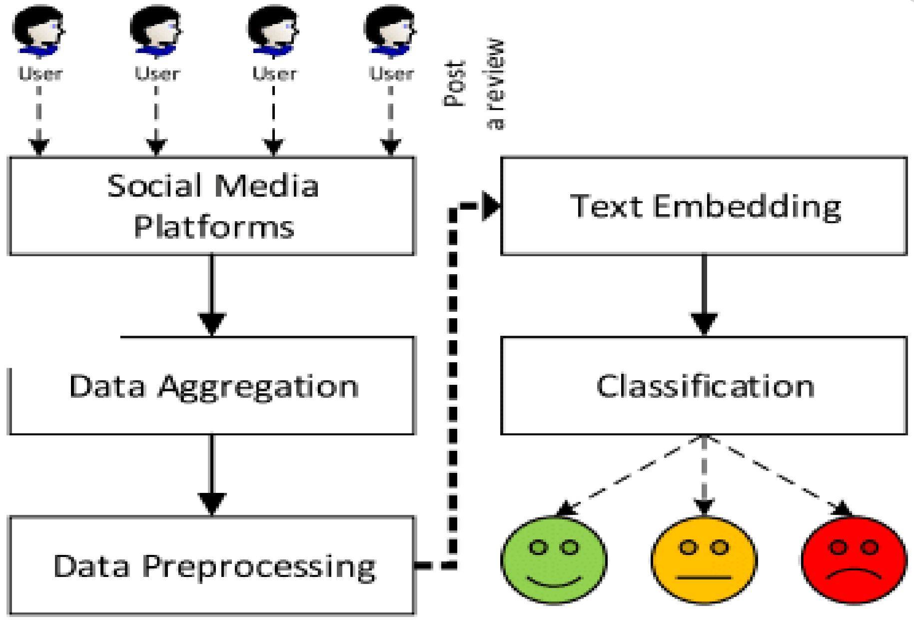
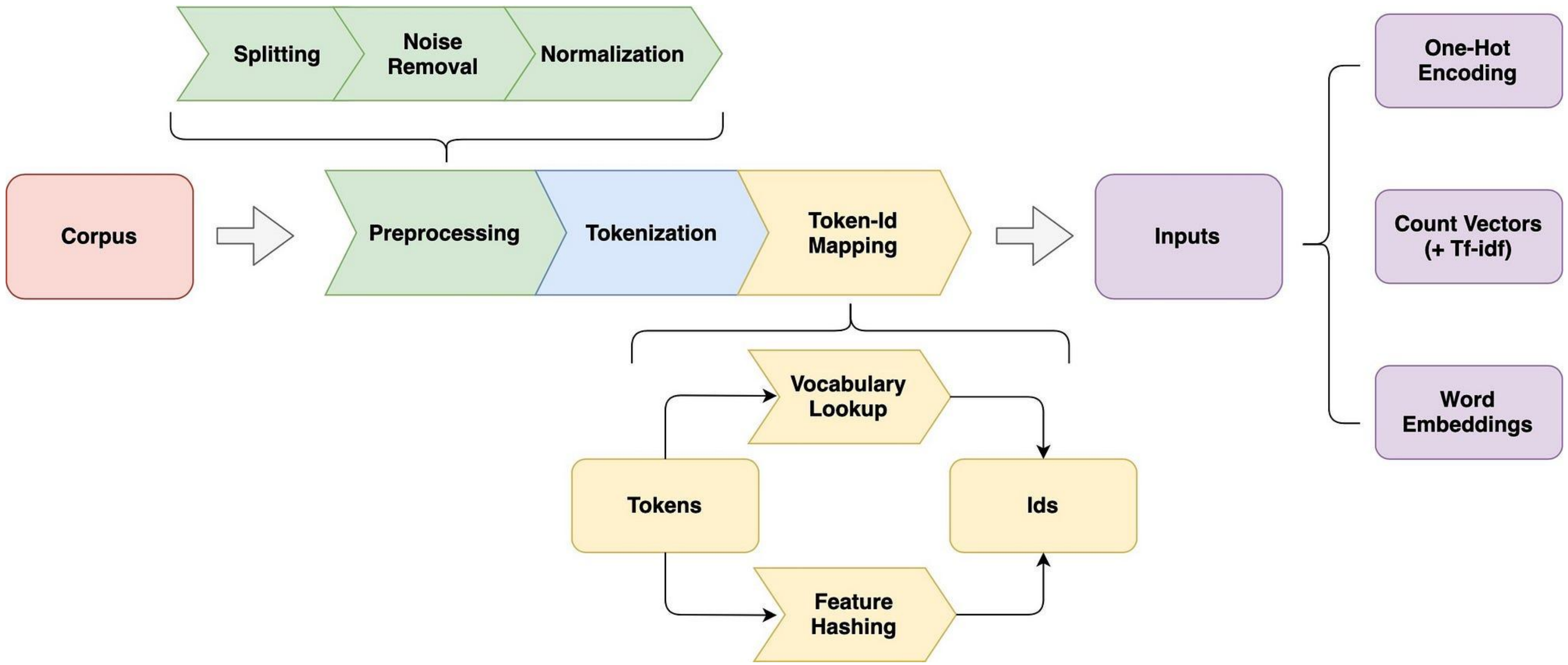


FIGURE 1. Basic steps of sentiment analysis on social media.

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:

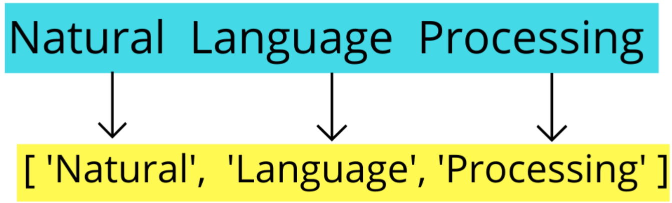
- Sentence: *"It was the best of times."*
- Tokens: ['it', 'was', 'the', 'best', 'of', 'times']

NLP Pipeline: From Raw Text to Insight

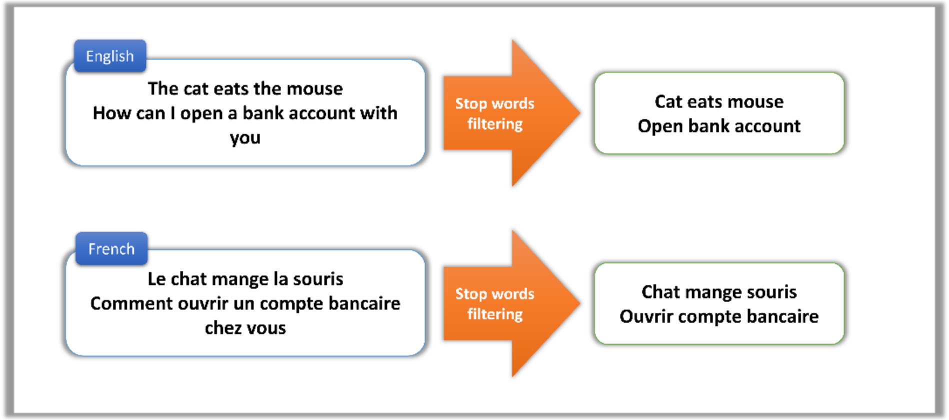
1. Text Preprocessing

- Essential to prepare raw language for modeling.

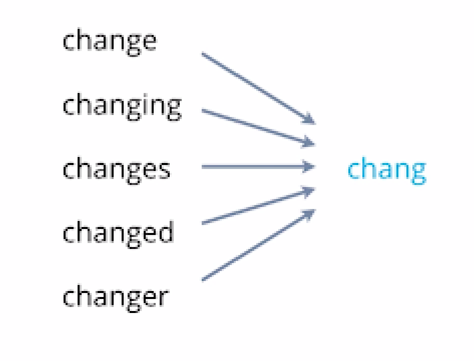
Tokenization



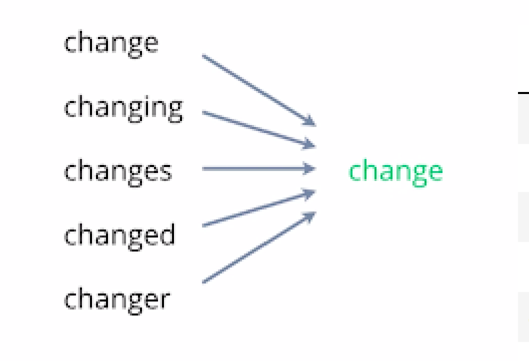
Removing Stop words



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:

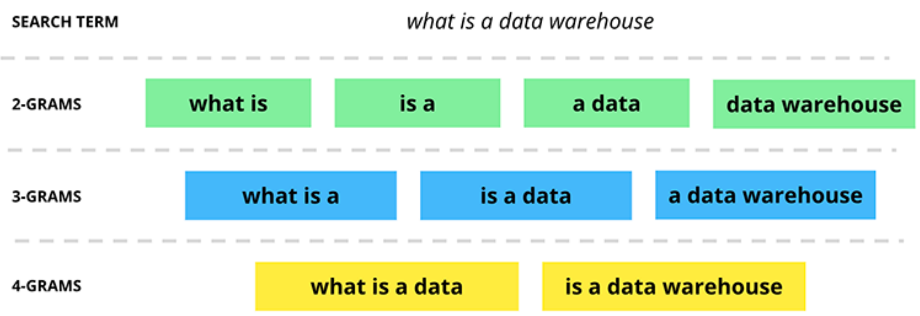
- Transform words or documents into numerical representations for machine learning models.

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]

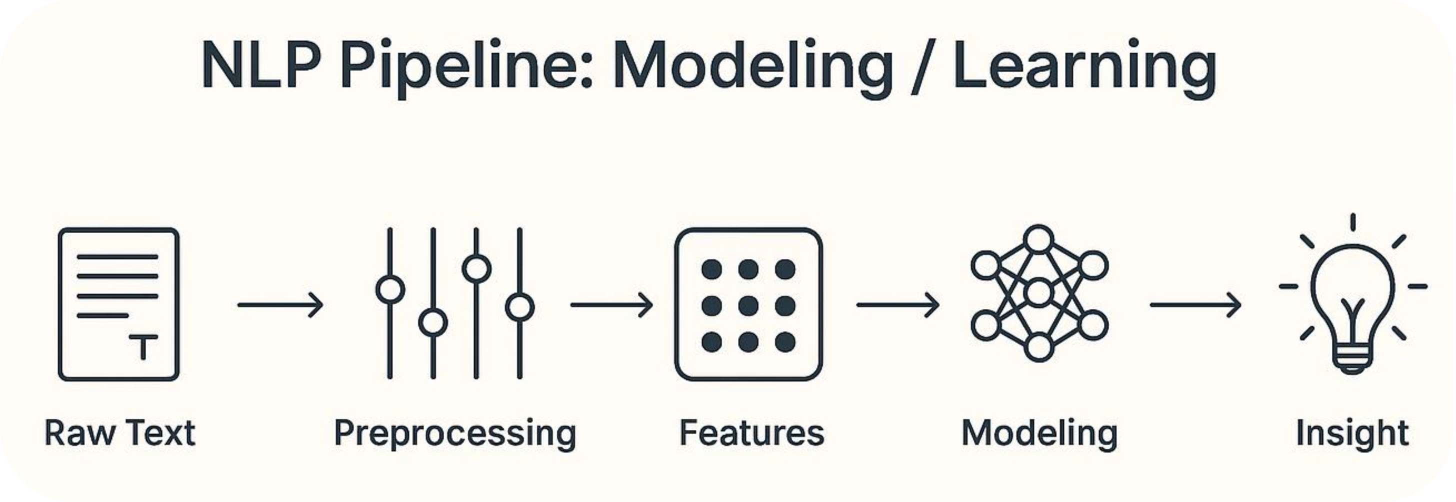
n-grams



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