Summer of Code

Artificial Intelligence

(Machine Learning & Deep Learning)

Instructor
Wajahat Ullah
- *Research Assistant* (DIP Lab)

Duration
03 Months
(September – November)

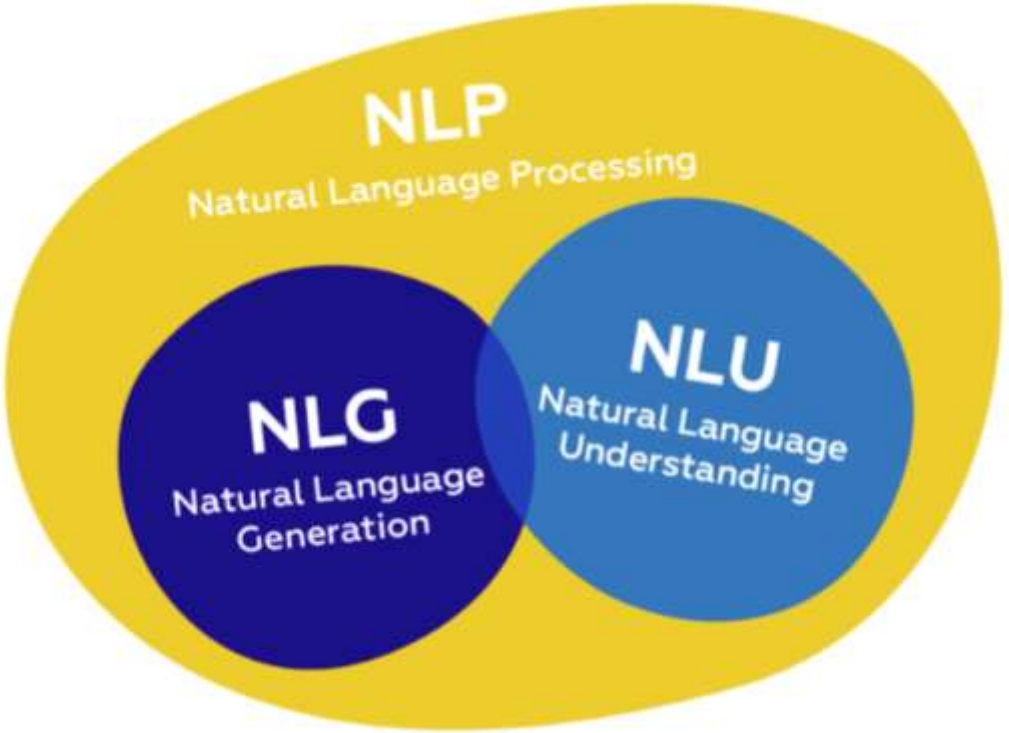
Day 03 – Deep Learning
(Natural Language Processing)

Objectives:

- Natural Language Processing
- Syntax, Semantics, Pragmatics, Discourse
- Data Preprocessing

Natural Language Processing (NLP)

- NLP is the discipline of building machines that can manipulate human language in the way that it is written, spoken, and organized
- NLP can be divided into two overlapping subfields:
  - **Natural Language Understanding (NLU)**
  - **Natural Language Generation (NLG)**
- It is separate from speech recognition.
- It is integral part of everyday life.



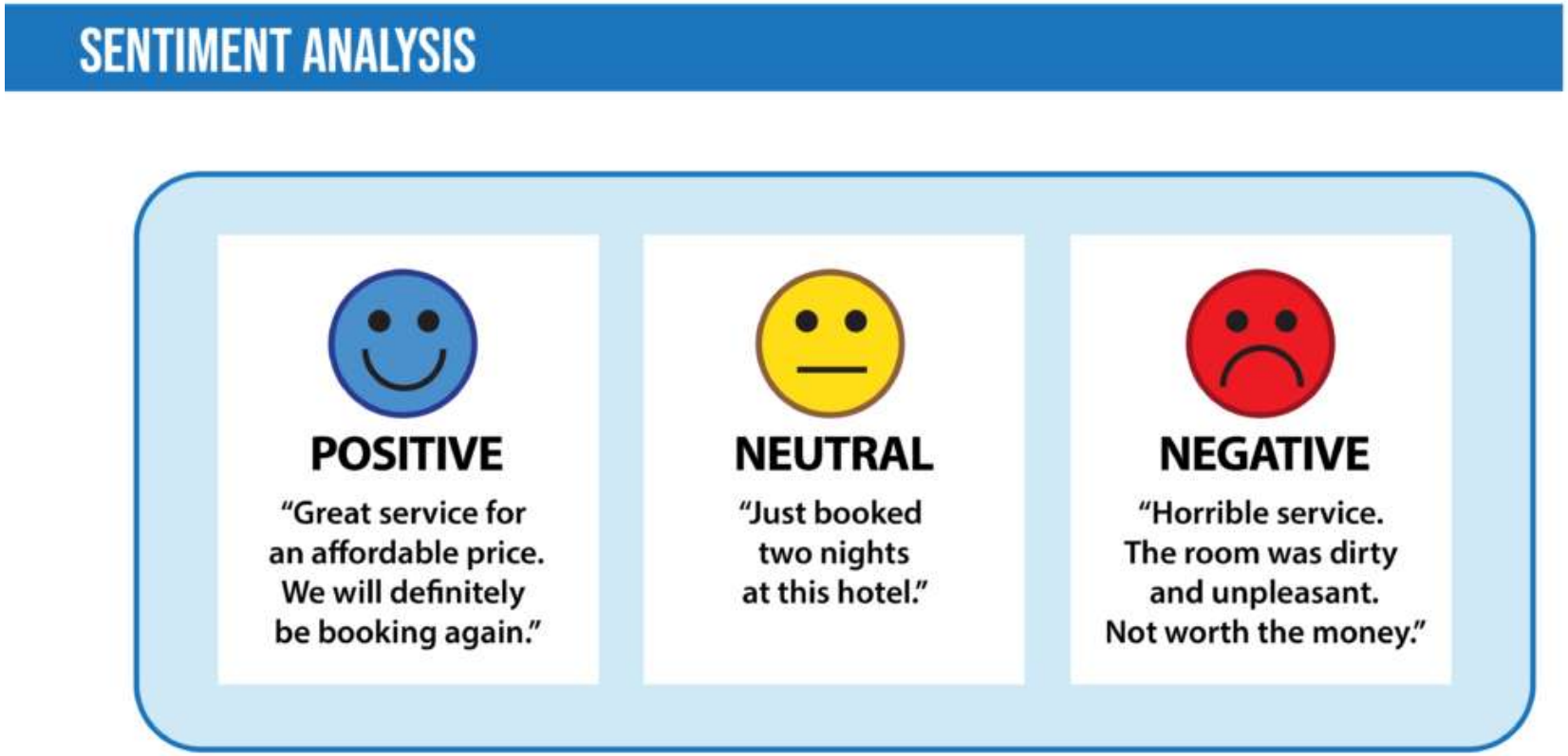
ChatGPT

Siri

Gemini

Applications of NLP

NLP is used for a wide variety of language-related tasks



Given text, sentiment analysis classifies its emotional quality.

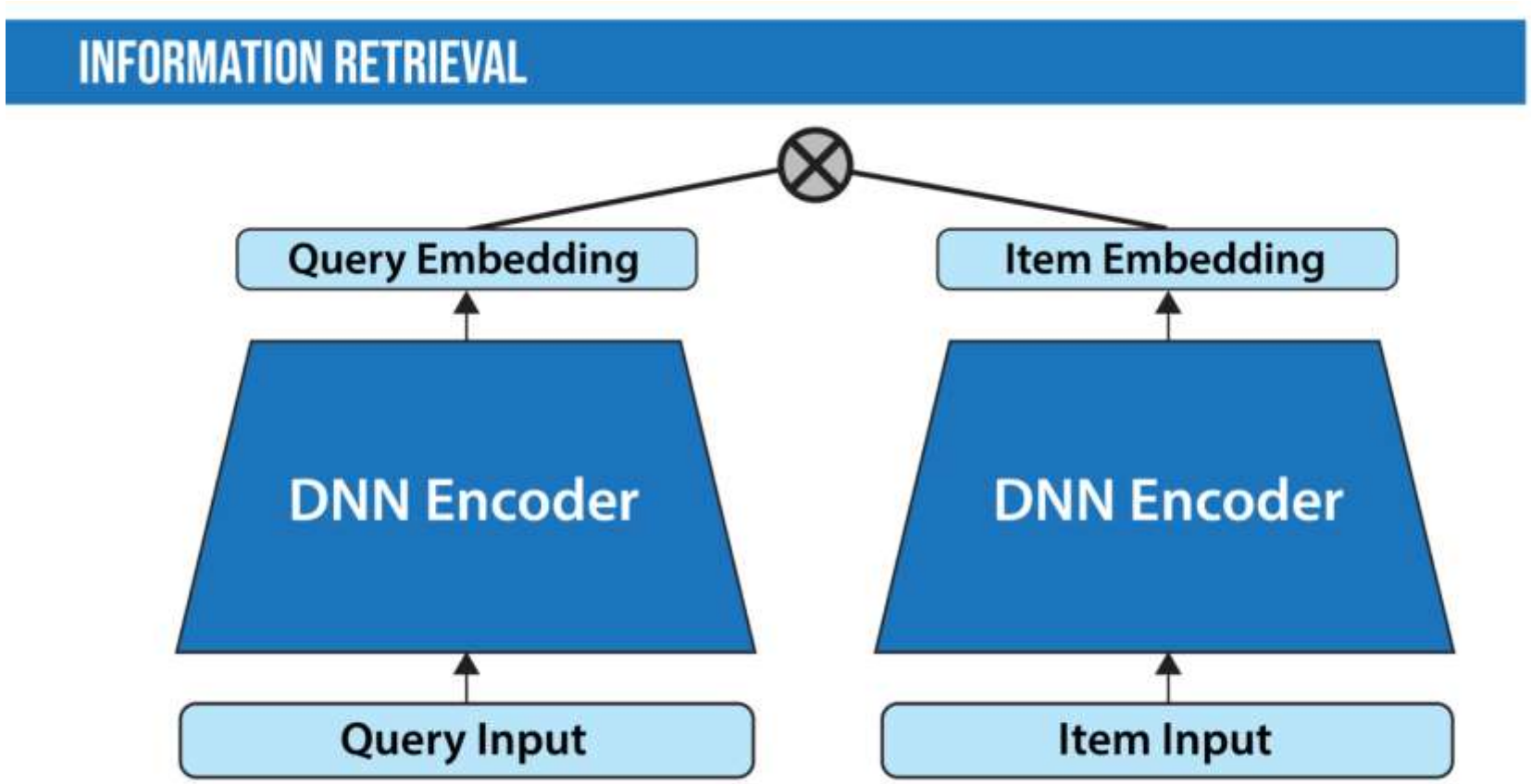
Applications of NLP

NLP is used for a wide variety of language-related tasks

spaCy named entity recognition tagging of the first paragraph of Andrew Ng's Wikipedia page. "NORP" stands for nationalities or religious or political groups. Note that spaCy incorrectly labels "AI" as "GPE," for geopolitical entity.

Applications of NLP

NLP is used for a wide variety of language-related tasks



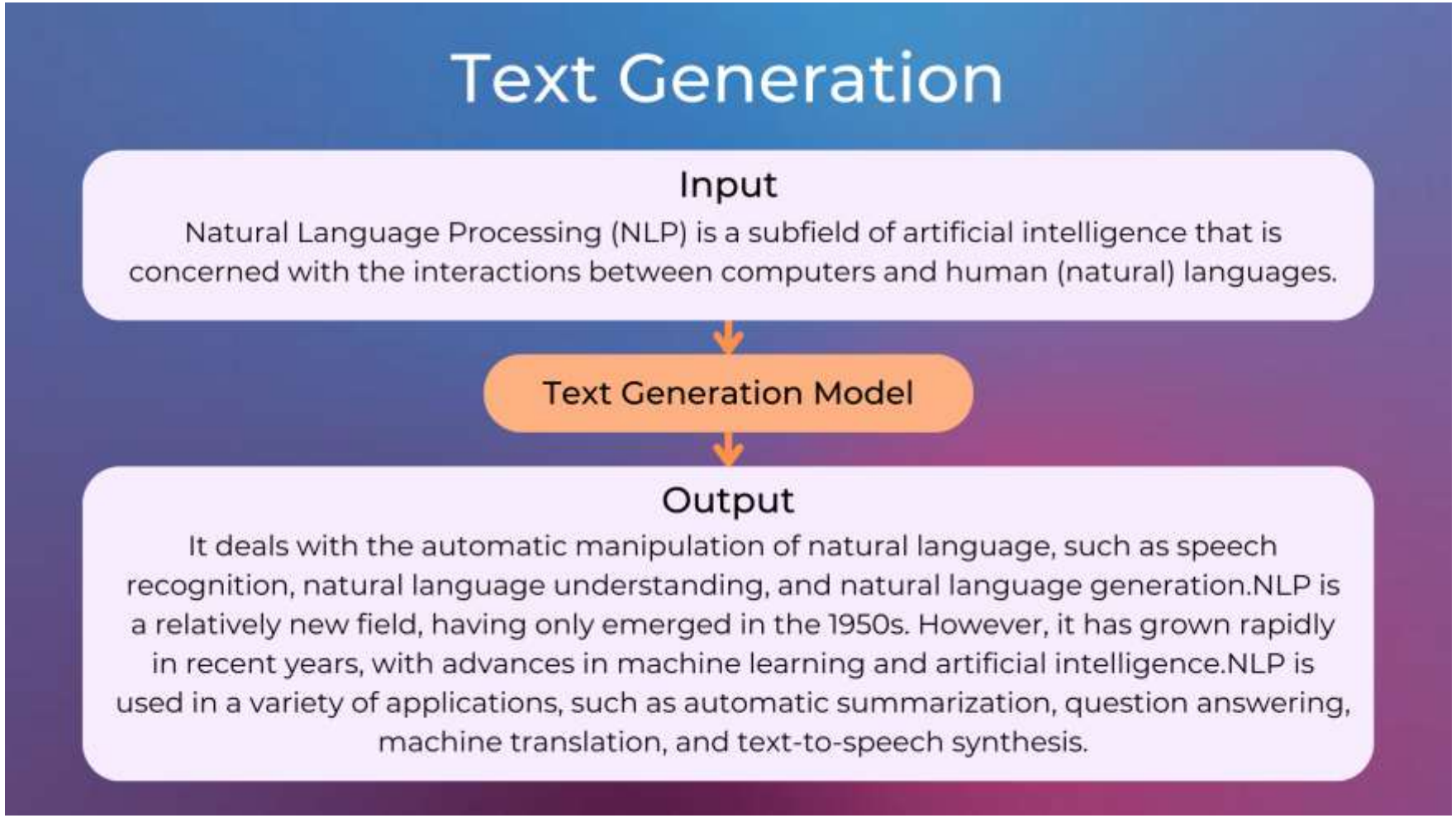
A two-tower network creates a representation of an input query and a group of documents (or items) through two separate networks. Then it compares the representation of the query with that of the documents to find documents that are most relevant to the query.

Applications of NLP

NLP is used for a wide variety of language-related tasks

Applications of NLP

NLP is used for a wide variety of language-related tasks



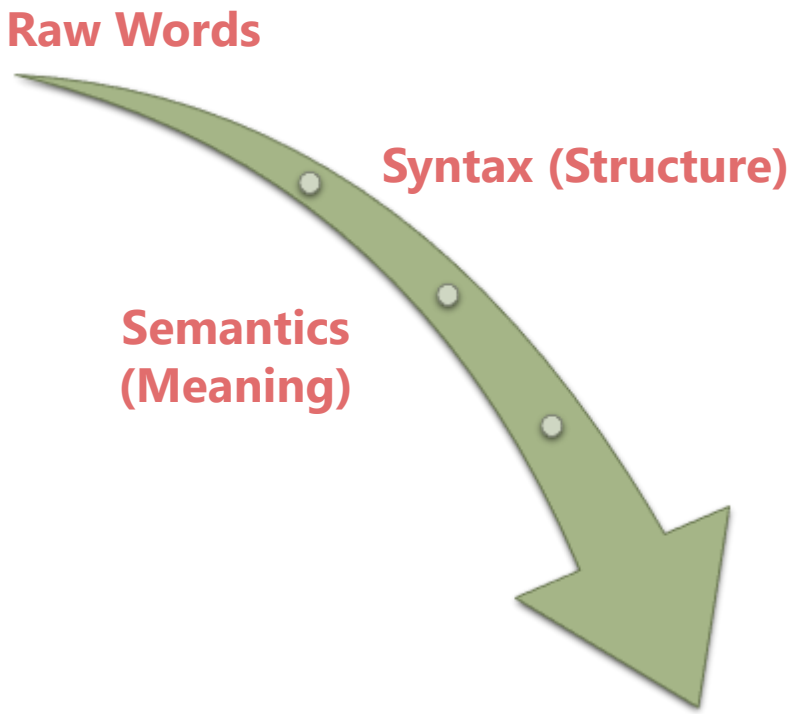
Understanding Language

Syntax (Structure)

- Rules that governs the structure of sentences (Grammar).
- It's about whether a sentence is correctly formed.
- NLP Task: Parsing (creating a parse tree to understand sentence structure).

Semantics (Meaning)

- Meaning derived from words, phrases, and sentences.
- What does this word or sentence mean?
- NLP Task: Word Sense Disambiguation.



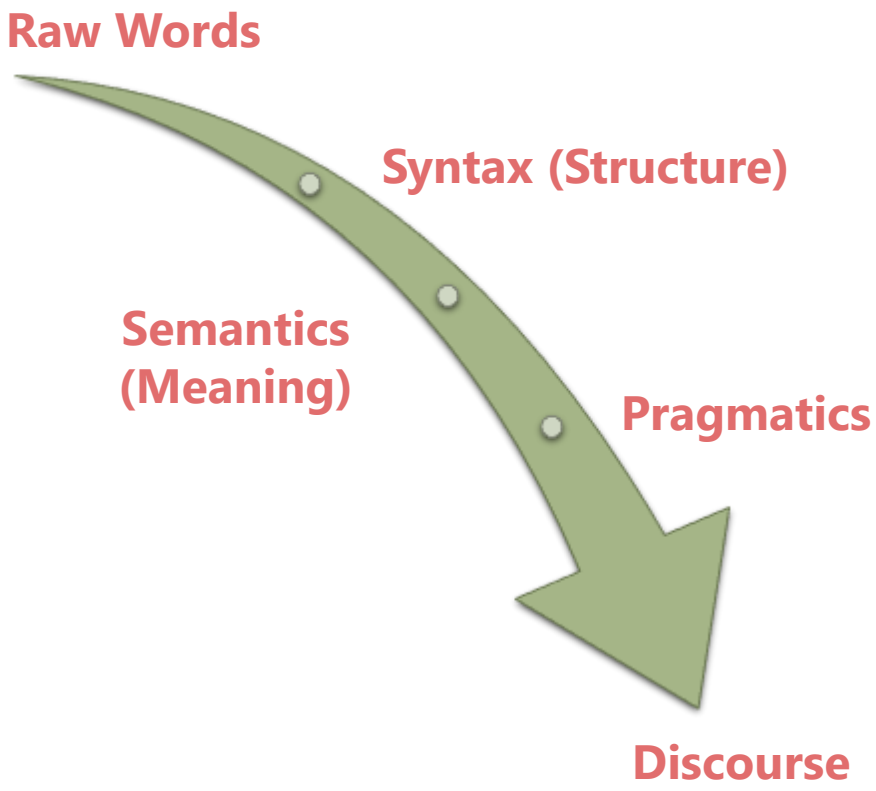
Understanding Language

Pragmatics (Context)

- How context contributes to meaning. It goes beyond the literal meaning to understand intent and implied meaning.
- The meaning depends on the situation, the speaker, and the listener.

Discourse

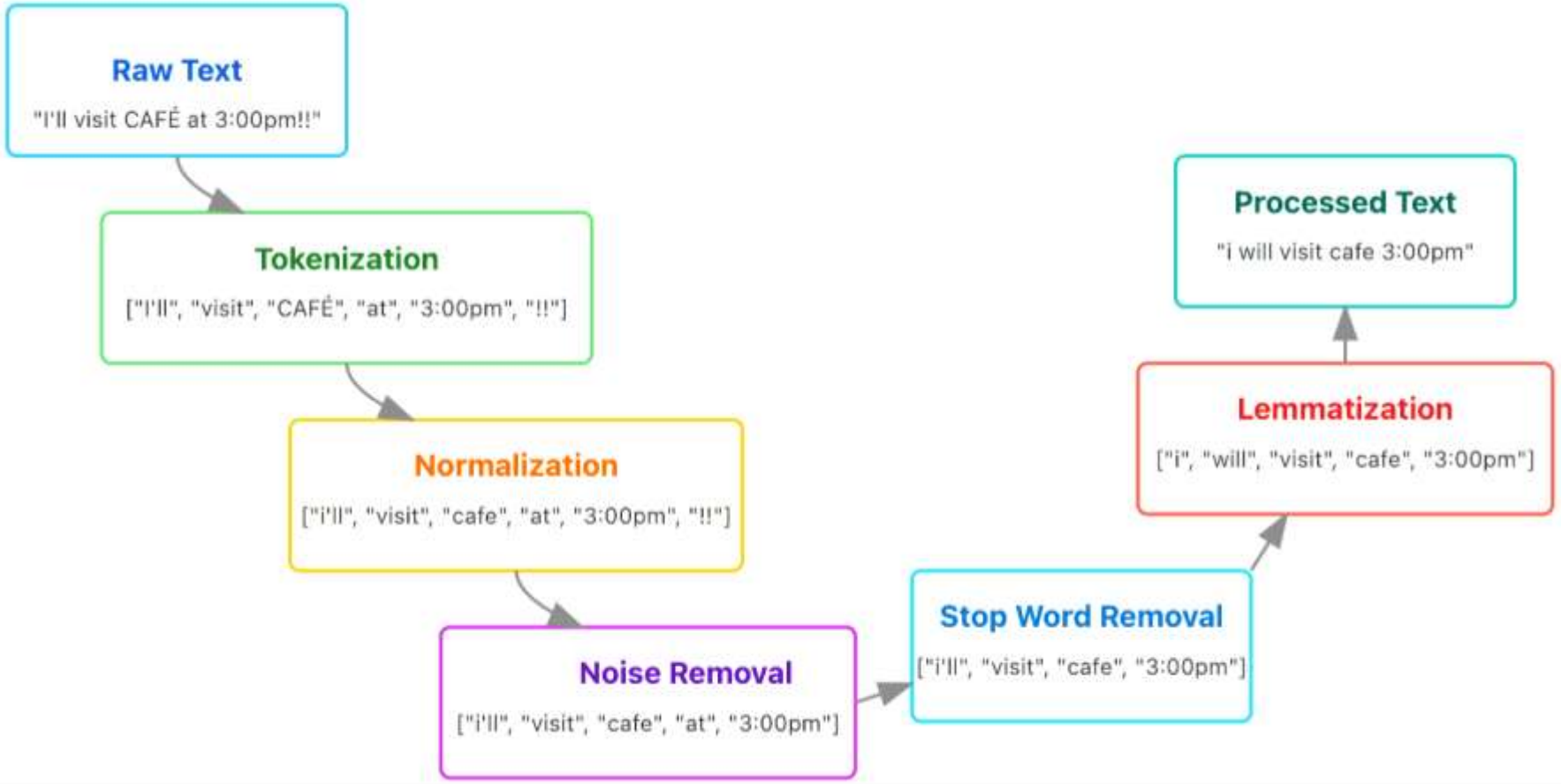
- Study of language beyond the single sentence. It looks at how sentences connect to form a coherent, meaningful whole.



How NLP Works?

Data Preprocessing

Raw Text is Messy! It's full of noise (punctuation, HTML tags), variations, and extremely common words that add little meaning



NLTK vs spaCy

Natural Language Toolkit

- Great for teaching, research, and trying out many algorithms. It's a toolkit.

spaCy

- Designed for real-world applications. It's fast, efficient, and provides a consistent API for a full pipeline.

FEATURE	SPACY	NLTK	HUGGING FACE
Pre-trained Models	Available for NLP tasks (NER, POS, etc.)	Limited pre-trained models	Extensive transformer models (BERT, GPT, T5, etc.)
Deep Learning Integration	Supports PyTorch and TensorFlow	No direct integration with DL frameworks	Seamless integration with PyTorch and TensorFlow
Multilingual Support	60+ languages (including multilingual models)	Limited, mainly English-based	Strong multilingual models (XLM-R, mBERT, etc.)
Ease of Use	Simple and production-focused API	Comprehensive but educational-focused	High-level API for deep learning models
Community & Ecosystem	Active community, production use	Strong academic community	Huge community, model hub, and collaboration
Best Use Case	Production NLP applications	Educational purposes and research	Cutting-edge NLP with transformers and fine-tuning

Thank You