

# Overview of Natural Language Processing (NLP)

*By: Shreya Valaboju*

**CS 4395.001**

Natural Language Processing or NLP is a branch of artificial intelligence (AI) that allows computers to process, analyze, and understand human language. NLP can be described as an intersection of linguistics, computer science, and artificial intelligence. AI is a more general concept and under its umbrella lies NLP. Today, many companies and other organizations invest large amounts of funding in NLP research and projects to leverage further and improve its usage. Some examples of modern NLP applications include chatbots, such as the infamous ChatGPT, customer service sentiment analysis, autocomplete functionalities, like search bars or text messages, and language translation. As part of defining NLP and its objectives, it is also important to note the distinction between two concepts in human-to-human dialogue: natural language understanding and natural language generation. Natural language understanding means that each party engaging in dialogue understands and comprehends what the other is saying. The latter, natural language generation, involves generating spoken responses or output. Both are important and used as a part of natural language processing.

The first main approach to NLP is rules-based. This approach is one of the first techniques used for NLP. It depends on applying a set of defined rules in order to process and understand human language. Some examples of this approach are converting words from plural to singular and using context-free grammar and regular expressions to generate rules for sentences. Although this technique is old, the rules-based approach is still widely used and can be effective for specific use cases. Some of those applications include spell-checking and simple chatbots, such as Eliza from the 1960s. However, depending solely on a rules-based approach is unreliable and unscalable to the many complexities of human language.

The second approach to NLP is more mathematical, specifically statistical and probabilistic. This method involves calculating probabilities and frequencies of words or phrases. Traditional machine learning models, like Naive Bayes, Logistic Regression, or Decision Trees follow this approach by learning statistical and probabilistic methods. For instance, in a probabilistic classifier, such as Naive Bayes, the likelihood and prior are estimated from the training phase. Although a very popular NLP technique, a modest amount of data and processing power is required. After more data became more accessible, statistical and probabilistic approaches started to take hold in the late 1980s. Forecasting or predicting text, sentiment analysis, and text summarization are common applications of this method.

The last approach to NLP is using deep learning. This technique requires a large of data and computing power. Because of advancements in cloud computing and GPUs, deep learning for NLP is slowly evolving and becoming popularized today. Common deep learning algorithms include recurrent neural networks, convolutional neural networks, and LSTMs (long-term short memory). These are more complex forms of basic, artificial neural networks. Deep learning is thought to have fixed many of the drawbacks of the rules-based and statistical and probabilistic

approaches to NLP. Its applications include language translation, generation, and understanding. An eventual goal of using deep learning for NLP is to get better at understanding context, remembering previous dialogue, and generating more human-like responses as possible.

As a computer science student, I have been exposed to various topics within artificial intelligence, including machine learning and NLP. I am deeply interested and curious in learning more in-depth about NLP because I believe it has a growing and significant role in future applications and technologies. Although I have worked on various NLP projects in past internships and courses, I want to further dive-in and expand my knowledge in this field. As an aspiring data scientist, I look to apply and practice the concepts learned from this course with expertise.

**Sources:**

- *Course Textbook (NLP with Python by Dr. Mazidi)*