

## Overview of NLP

Hrithik Ochani

a. NLP (Natural Language Processing) is the systemic comprehension of natural (human) languages, mostly done by computers. Computers break down natural languages into rules that they understand and evaluate syntax and meaning afterwards.

b. AI benefits greatly from NLP, especially pertaining to communication between AI and humans. AI comes to understand more of human language and conversation through NLP, learning what is common, sentence structure, etiquette, and so on. AI and human communication is a high-level application of the results of NLP.

c. Natural language understanding is receiving an input of natural language and being able to derive meaning from it. Natural language generation is knowing the syntactic rules of natural language and creating a sentence that is functionally correct, but not necessarily knowing of its meaning. These two ideas can loop with one another to achieve greater generation and understanding.

d. Spell-checkers, chatbots, search auto-fill, voice generation (or recreation), and lip-reading conversion to sentences.

e. There are three main approaches to NLP: rule-based, statistical, and deep learning. Rule-based NLP takes basic definitions and structuring (creating rules) of human language and attempts to understand and convey language through these rules. For example, sentences have verbs and nouns, and there are rules for their placement and interaction. Unfortunately, this approach doesn't scale well with the actual complexity of human language. Examples of rule-based NLP applications include spell-checkers and context-free grammars.

Statistical NLP takes into account the frequency and usage of words and orderings to try and create a more accurate representation of human language. It uses traditional machine learning algorithms, but requires a lot of input data and a good amount of processing power. Examples of statistical NLP applications include suggesting auto-fills and summarizing text.

Deep learning NLP is an expansion of statistical NLP that had improved results in language transition, language generation, language understanding, and much more. However, it takes an incredibly large amount of data, as well as extremely strong processing power. It's the new, trendy option that may be overdriven by hype. An example would be the highly-advanced AI of today.

f. My personal interest in NLP comes from an intrigue in rule-based language processing, a concept I encountered in both Paradigms of Programming Languages and Automata Theory. Learning to work with context-free grammars and creating parse trees of sentences (or more commonly, expressions) were interesting experiences that I wanted to expand on. I understand

that NLP, especially its modern applications, is usually beyond rule-based processing, but I am still interested in the statistical approach and how it helps develop AI.