

What is Natural Language Processing (NLP)

- Closely related to the fields of computational linguistics and machine learning
 - Linguistics is concerned with language itself; its structure and how we represent it, the nature of meaning derived from it, the application of logic and reasoning over it etc..
 - In contrast, NLP is aimed at building computational methods that enable computers to process human language with the goal of performing some task. Examples include text search, question answering, conversing with humans etc..
 - NLP relies heavily on various machine learning methods, for example deep neural networks, but uses them as a means to accomplish specific tasks. However, many novel ideas and progress in ML have come from NLP, for example transformers (BERT); NLP and ML have a symbiotic relationship.

Machine Learning

- At a high level, machine learning is set of techniques, or *machinery*, that is designed to automatically learn complex relationships between random variables. Typically there is some output that we are interested in predicting, and there are inputs that we hypothesize affect that output. The goal is to build a model that accurately predicts the output given the input.
- In the context of natural language there are several things that make it unique:
 - The **distribution of words is highly non uniform** in the wild and follows a power law with a very long tail [1].
 - Unlike images or tabular data, **language is highly compositional** and meaning is derived from how characters, words, and phrases are composed. There are countless ways of representing the same meaning using language; this makes it hard to automate the extraction of meaning from it.