

1. Natural Language Processing (NLP) is focused around implementing the complex language recognition and processing abilities of the human brain into computers. This process is automated and can then be scaled up by several orders of magnitude in order to process large sets of text data.

2. NLP is an application/field of Artificial Intelligence (AI) which is used to process text data similarly to the way a human might. The idea of AI is to create machines which can function with an “intelligence” modeled after humans, and NLP is a specific intelligence related to language understanding.

3. Natural Language Understanding (NLU) is the processing of textual information while Natural Language Generation (NLG) is the creation of textual information. While these two are distinctly different processes, they are closely related to each other. NLG can be done to generate a response based on NLU processing a response and vice versa.

4. Rules-based approaches: These are the oldest techniques for NLP which follow specific grammar-based rules using regular expressions and exceptions. Some examples of this would be expanding contractions or scanning text for greetings and responding with a greeting. This approach can be inefficient because it is extremely difficult to scale it up in such a way that it could handle all the complexities of different languages. However, it can still be an effective tool when used to handle simple text-based challenges, such as video game dialog.

Statistical and probabilistic approaches: Instead of creating rules for particular scenarios when processing text, statistical and probabilistic approaches apply mathematics to better understand textual data. By calculating the probability of words and sequences of words in different contexts, a language model can make more accurate assumptions on their usages and meanings. For example, the term “hot sauce” could be recognized as “sauce that is hot” or “sauce that is spicy” depending on the context of the whole sentence and a probabilistic approach could accurately assess this. Some statistical and probabilistic algorithms are Naïve Bayes, Logistic Regression, SVMs, Decision Trees, and small Neural Networks.

Deep learning: Deep learning developed from small neural networks, such as the ones used in statistical and probabilistic approaches, to attempt to model the complexity of human neural networks by implementing multiple neural layers. This NLP approach is uniquely powerful in the fact that it does not require extensive knowledge on the data prior to training. It can self-adapt, find new patterns, and process massive amounts of data. One example is the use of deep learning to adapt to different accents of a particular language.

5. I’ve recently become fascinated with linguistics and the way languages can be broken down and understood at a fundamental level. NLP is the perfect convergence of my passion for Computer Science and linguistics, and I think it is extremely interesting how computers are quickly being made to mimic many of the complexities and nuances of the human mind when it comes to human language interactions. I would love to work on a live translation device for people to be able to communicate with any person at any given time, and NLP would be a huge part of this project.