

Overview of NLP

a. Natural language processing, or NLP, is the area of computer science concerned with computers being able to make sense of and interpret natural human language as we speak and write it.

b. NLP is the subsection of AI that has to do with understanding natural human language

c. Natural language understanding, or NLU, is the ability of the computer to understand the meaning of natural human language. Natural language generation, or NLG, is the conversion of the structured data from NLU into sentences that humans can more easily comprehend.

d. Examples of modern NLP applications include smart assistants (Siri, Alexa, etc.), autocorrect, and online language translators.

e. The first main approach of NLP is rules-based approaches. With this approach, there is usually more focus on pattern-matching with things like regular expressions. This approach can also be used to check if a particular sentence is grammatically correct with things like context-free grammars. Rules-based approaches are known as the oldest methods used in NLP. They are also known for being difficult to scale up since human language can be very complex and it's hard to fully capture human language with rules alone.

The second main approach of NLP is statistical and probabilistic approaches. These approaches can form language models using a total count of words as well as the probability of words and sequences of words appearing. This idea can be seen in practice with things like predictive text and autocorrect to help the user fill in the blank with what is most likely to be typed. Machine learning algorithms such as Decision Trees and SVMs fall into this second main approach as well since they also use statistics and probability for their learning. The issue with these approaches is that you need the necessary processing power and at least a moderate chunk of data to work with.

The third main approach of NLP is deep learning. This approach involves a computer learning from large amounts of data, similar to how human beings tend to learn by example. Some deep learning algorithms include convolutional neural networks and recurrent neural networks. One issue with this approach is that you need a lot of data (petabytes) and the capability of processing that much data, which not everyone can afford and/or access. This means that NLP applications today are using more small-scale

deep learning techniques. Given that this approach is the more “cutting edge” of the main three, there is a lot of hype surrounding it as well.

f. NLP is a subject that I don’t know a whole lot about, but it seems interesting. Also, for this course in particular, I read that we would be using Python, which is a language that is very popular right now, so I figured it would be wise to get some more experience with using it since I haven’t used it that much up to this point. I’m not entirely sure what I would like to do yet once I graduate from UTD, so hopefully this class and this subject matter can help guide my decision and help make that process a bit easier for me.