Cady Baltz
CS 4395.001 Human Language Technologies
Portfolio Assignment 1: Overview of NLP

a. **Define NLP in your own words**

Natural language processing is a method of teaching a computer how to understand human languages and how to generate its own responses in the language.

b. **Describe the relationship between NLP and AI**

NLP is a branch within the field of AI (Artificial Intelligence). Similarly, ML (Machine Learning) is another branch of AI that is often applied to NLP. NLP projects can use any combination of NLP techniques, ML techniques, and general AI techniques. These fields are all rapidly developing and continually influence each other.

c. **Compare and contrast natural language understanding and natural language generation**

Natural language understanding means that someone (or a computer) is able to understand text produced by someone else, while natural language generation is the ability to produce text independently. Both of these processes are essential for successful conversations between two people or computers.

d. **List some examples of modern NLP applications**

- ChatGPT
- Google Translate
- Chatbots
- Autocorrect
- Email spam detection
- Voice assistants
- Sentiment analysis

e. Describe each of the 3 main approaches to NLP, and list examples of each approach

1. Rules-Based Approach

Rules-based approaches use a set of defined rules to process natural language. For example, a rules-based approach would process text to convert plural words to singular words by removing the -s at the end of words, and would need to include specific rules for exceptions like the word 'children'. The main problem with this approach is that human language is very complex and does not have a context-free grammar, making it impossible to define a complete set of rules for processing it. The main benefit of this approach is that it is a relatively fast and easy approach for processing simple texts.

Examples of Rules-Based Approaches:

- Eliza
    - Eliza was a virtual therapist created in the 1960s which used regular expressions to respond to user input.
- Text extraction
    - Rules can be defined to extract specific elements from a body of text, such as phone numbers.

2. Statistical and Probabilistic Approach

The statistical and probabilistic approach to NLP involves using math to process text. For example, a simple approach in this category is counting words to find the probability of certain words and sequences of words. This approach is frequently used with ML, as many ML algorithms apply statistical and probabilistic methods, such as Naive Bayes, Logistic Regression, and Decision Trees. The main benefit of this approach is that it works well on moderate to large datasets, requiring less data than a deep learning approach while still allowing for more complexity than a rules-based approach.

Examples of Statistical and Probabilistic Approaches:

1. Predictive text
    a. This approach is used to predict what a user will type next, such as autocompleting a query in a search bar.
2. Language translation
    a. This approach is useful for translating between different languages as it can help choose between words with similar meanings in the resulting language.

3. Deep Learning Approach

The deep learning approach has been made possible as large language datasets are now more widely available. This approach involves algorithms such as recurrent neural networks, convolutional neural networks, and LSTMs, which all evolved from the basic neural network algorithm. The main problem with this approach is that it requires massive amounts of data, as well as expensive hardware (e.g. GPUs) that can handle all of this data. The main benefit of this approach is that it can handle more complex interactions by, for example, considering the context of a text and remembering previous responses in a conversation.

Examples of Deep Learning Approaches:

1. GPT-3
    a. GPT-3 is a language model that uses deep learning to generate human-like text.
2. Conversational chatbots
    a. Modern chatbots use deep learning to handle complex interactions with human users.

f. Describe your personal interest in NLP and whether/how you would like to learn more about NLP for personal projects and/or professional application

I am interested in NLP as a result of my interest in language in general. I have always loved my foreign language classes, including Spanish, French, and Mandarin, and I still love using Duolingo. I also competed in spelling bees when I was younger and found it really interesting to study the roots of words and how they evolved from different languages. NLP is a great way for me to combine my interest in language with my interest in computer science. I had the opportunity to intern on the Siri team at Apple where I gained exposure to natural language processing and AI at a professional level, but I have never actually learned the theory behind this technology. By taking this class, I hope to gain a stronger understanding of modern approaches to NLP, and how I can practically apply them in future personal and professional projects. I am excited that this is a project-based course, as I learn concepts best by being hands-on.