Overview of ML

What is Machine Learning:

Machine Learning is the leveraging of algorithms and models to find patterns from datasets with any amount of data.

The importance of data in Machine Learning:

The importance of data cannot be understated when it comes to Machine Learning. This is because any pattern that an algorithm may find is based off of the data that is already present. Whether we want to extrapolate future results, discover a relationship between data, or generate a result for some new data. If our training set is flawed or not properly cleaned, we can have any number of inaccuracies in our algorithm. Thus, when designing data, we have to ensure we account for any of the many factors that could negatively affect the accuracy of our resulting code.

The relationship between AI and ML:

AI is simply ML that is attempting to replicate the human brain in some form whether that be neural networks or other models. ML is the data and algorithms that AI is built upon.

Examples of Machine Learning Applications:

1) Stock Trading Pattern Detection
   We find these patterns using large datasets. In traditional programming we can only look for patterns that we are already aware of, but by analyzing large amounts of data new patterns can be generated.
2) Image recognition

   We can't use traditional programming when we want to recognize items from a picture because there are many angles and differences between a single item among different pictures. This means that we can't program to look for some preset number and color of pixels because no two images are the same. We require many different images of a single object for an algorithm to recognize similarities between a single object.

Define:

When we import a set of data into a data frame we have many rows that each define a single observation. We use these individual observations to attempt to predict patterns from them. Inside of a single observation we can have many individual features that define that observation from either a qualitative or quantitative perspective. Qualitative is usually classification of some type. For example, the color that an observation is such as Red, Blue, etc. We can use numbers to index the qualitative data saying red is defined as 1 and blue is defined as 2. This is helpful from the algorithms perspective since

character or strings don't mean much. When we use quantitative data we are usually defining something that can't be sorted into exact categories and has a much higher range of values. We can use these quantitative data points to define with much greater accuracy the value that made that number.

Personal Interest:

Personally I have already done projects in classification and NLP, but I am most interested in working with reinforcement learning. From a professional stand point one of the things that made ChatGPT so interesting is that it not only used the GPT engine, but it had thousands of hours of being trained by real people using reinforcement learning techniques. This allowed people to guide the algorithms approach to what a proper response would be when asked a question. I want to explore other potential applications of reinforcement learning.