Konrad Heyen & Mason LaFerney

October 9, 2020

Dr. Baird

AI Project Proposal

No one ever reads the terms of services when opening a new website or buying a new product. There are many problems with this phenomenon, namely data tracking and the loss of expected rights when using a company’s website or product. The aim of this project is to create a Terms and Services summarization application using artificial intelligence to determine important information contained in a document. The program will make use of deep learning networks to process a wide variety of textual inputs and their summaries and then predict the appropriate summarization of a given document.

There is currently no artificial intelligence with the same goal in mind, however, there is a chrome extension named *Terms and Services; Didn’t Read(a)* which summarizes the terms and conditions of certain companies. This summarization and categorization is done manually by a team of lawyers. Multiple AIs capable of summarizing textual information already exist. One of the most successful, PEGASUS(b), is an abstractive text summarizer that makes use of sequence to sequence (seq2seq) learning. PEGASUS uses transformer encoder-decoder models combined with self-supervised pre-training in order to model the text summarization task.

In the context of the project, language will be processed and generated in a symbolic form, similar to PEGASUS. Python will be the primary language used in this project due to its extensive neural network libraries and capabilities. The goal of this project is to learn the fundamentals of artificial intelligence and specifically gaining knowledge in the creation and application of neural networks. Our intended product does not significantly differ in output from the work that is accomplished by the team of lawyers from *Terms and Services; Didn’t Read,* however, the method used to achieve the output will require no manual labor*.* The final working version of the application is intended to be completed by the week of December 7th and will then be presented to the class in the same week and distributed on GitHub.