DSO 560 – *16305: Text Analytics + Natural Language*

Professor: *Ash Pahwa*

Email: ashpahwa@marshall.usc.edu

When: *Spring 2024, Second half of the semester. Thursday 6:30 – 9:30 PM*

Office: *BRI-303E*   **Units:** *1.5*

WHY TAKE THIS COURSE?

Language models (ChatGPT, GPT-4 etc.) use deep learning techniques to identify patterns and relationships in language data, allowing them to recognize and generate human-like text. One of the key advantages of large language models is their ability to learn from context. Hardly a day goes by when applications of Language Models do not appear in the national news. All technology companies like Google, Microsoft, Amazon are actively hiring people with NLP background.

COURSE OBJECTIVES

- Analysis of Text Data
- Study the design and architecture of Language Models like ChatGPT. GPT-4
- Study the procedure of generating Word Embeddings
- Analyze Transformer architecture – encoder & decoder. Encoder/Decoder architecture is the foundation of Language Models

KEY CONCEPTS

- Regular Expressions
- Tokenization/Vectorization of Text
- Semantic Analytics
- Text Classification
- Neural Networks and Deep Learning
- Word Embeddings
- Transformers
- Language Models – ChatGPT, Bard

COURSE DESCRIPTION

Recently Language Models based chat boxes like ChatGPT (OpenAI), GPT-4 (OpenAI), Bard (Google) are in the news. These Chat boxes are the foundation of Artificial Intelligence. All major tech companies are moving towards dominating this new industry.   The foundation of these Language models is Natural Language Processing.

The main goal of NLP is to understand the meaning of text. Only when computers understand the real meaning of the text, can they take decisive action which must be the intended action. Sentiment analysis of text is one of the important applications of NLP. The use case of sentiment analysis is for the purpose of analyzing customer feedback and tweets. The translation of text between languages is another significant NLP application.

There are currently two different approaches to NLP. The first one is the analysis of words, sentences, and the semantics of text. There are various software packages that can provide these capabilities. These software packages are Natural Language Tool Kit (NLTK), TextBlob, and spaCy. The other approach to NLP is using the Machine Learning strategy to analyze the text. Neural Network models are used to train a model by feeding it a lot of text data. Both approached will be covered.