# NLP, NER and POS Tagging

Michael H. Shimanyula

University of San Diego

Masters in Applied Artificial Intelligence

# Abstract

In recent times, claims related to climate change have been growing in both scope and number. Verifying the validity of these claims has never been more critical. This study uses Natural Language Processing (NLP) techniques, alongside Amazon Comprehend, to analyze a dataset containing various climate-related claims. We focus on different aspects of text analytics, including tokenization, part-of-speech tagging, named entity recognition, sentiment analysis, and key phrase extraction.

# Methodology

**Data Preprocessing**

The initial step involved reading a JSONL file, named climate-fever.jsonl, containing climate-related claims. The Python library SpaCy was used for NLP tasks. We used the following code snippet to extract critical information from each claim.

**Tokenization, Lemmatization, and Stop Words Removal**

For each claim, the text was tokenized, lemmatized, and stop words were removed using the following SpaCy commands.

**Results**

POS Tagging: Most of the claims include a high frequency of nouns, indicating that they often discuss specific entities or subjects.

Named Entities: Locations and organizations are frequently mentioned entities, underscoring the global and institutional scope of the climate discussion.

Sentiment Analysis: The majority of claims had a neutral sentiment, suggesting that the dataset is primarily factual or informative.

Key Phrases: Phrases like "climate change," "global warming," and "carbon emissions" appeared frequently, marking them as central themes in the discourse.

**Challenges Faced**

One of the main hurdles we encountered was that some parts of our dataset didn't have information about specific names, places, or organizations that the text was referring to. This type of information is usually identified through a process called Named Entity Recognition (NER). Not having this made it harder to fully understand what each piece of text was talking about.

# Conclusion

The study showed that text analytics could offer significant insights into the nature and scope of climate-related claims. Further work with a more extensive dataset and additional analytical methods is necessary for more comprehensive findings.

# References

Amazon Web Services, Inc. (n.d.). Amazon Comprehend – Natural Language Processing (NLP) and Machine Learning (ML). Retrieved from https://aws.amazon.com/comprehend/