Final Project: Proposal

#### **Topic:**

***A machine learning driven image caption generator.***

### **1. List 3 questions that you intend to answer (1 point)**

1. What features can and will be extracted from the images?

2. Which word embedding method will be optimal for this task (One Hot Encoding, Word2Vec, FastText, etc.)?

3. Which biases are inherent to the dataset - will our model be better at captioning animate or inanimate objects?

### **2. List all the datasets you intend to use (1 point)**

* Flickr30 dataset from the Grainger College of Engineering (University of Illinois)

<https://forms.illinois.edu/sec/229675>

* Microsoft COCO: Common Objects in Context (this dataset is something we may or may not use depending on the scale of our solution): <https://cocodataset.org/#home>

### **3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)**

To truly answer the questions posed above, we need to develop a skeleton model for our project. In order to build an effective image captioning algorithm we need to tackle a few specific tasks:

* Data Collection: We intend to use the datasets mentioned above for our model training.
* Data Exploration: Exploratory data analysis can be done to understand the dataset's structure and characteristics. This involves analyzing the distribution of image and caption lengths, identifying missing or inconsistent data, and visualizing the dataset to determine any patterns or trends. This step will be key to feature extraction and understanding the different elements of the image feature space. It will also be important in understanding the inherent biases in the dataset and seeing if our model will be better at captioning animate vs inanimate objects.
* Data Cleaning: Data cleaning is essential to remove noise and inconsistencies in the dataset. This involves removing any duplicate data, standardizing the format of captions, and removing any irrelevant data. Additionally, pre-processing techniques such as tokenization, stemming, and lemmatization can be applied to the captions to reduce the dimensionality of the data.
* Data Integration: In this case, data integration might not be necessary as we are primarily aiming to use the Flickr30 dataset. However, we integrate the Microsoft COCO dataset, data integration techniques such as data merging or concatenation may be required.
* Data Analysis: Deep learning techniques such as convolutional neural networks (CNN) and a Long Short Term Memory Network (LSTM) can be used for image caption generation. CNN can be used to extract the features from the image, and LSTM can be used to generate captions by predicting the next word in the sequence given the previous words. Evaluation metrics such as BLEU (bilingual evaluation understudy) score and METEOR (Metric for Evaluation of Translation with Explicit ORdering) can be used to evaluate the quality of the generated captions. We can study the efficacy and use of different word embeddings at this stage.
* Data Product: The data product for this project can be an interactive web app or a report that generates captions for the uploaded images. The app can be designed using a web framework like Flask or Django, and the model can be deployed using cloud platforms like AWS or Google Cloud. The user can upload an image, and the model will generate the caption for the image, which can be displayed on the app.

### **4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)**

1. Captioning for visually impaired (translated caption to Braille). Image caption generators can provide textual descriptions of images, making visual content accessible to people who are blind or have low vision. This can help to create a more inclusive and equitable society.

**For Future Use:**

Image search optimization: A caption-generating model can be used to generate captions for images that are used in search engine optimization (SEO). By providing descriptive captions for images, search engines can better understand the content of the image and provide more relevant search results.

Social media content generation: A caption-generating model can be used to automatically generate captions for social media posts. This can save time and effort for social media managers and influencers, while also ensuring that captions are descriptive and engaging.