

CSCI 6883: Master's Degree Project

Answer the questions below using at least two and no more than three pages. Your final document must adhere to having,

- Single spaced lines.
- Letter sized (8.5" by 11") or A4 paper and an inch of margin on all sides.
- 12pt font size along with the font family either being Latin Modern or Times New Roman.

Question 1. Give an [elevator pitch](#) (no more than 100 words) for your final project idea.

"My project is an image captioning system that uses deep learning to generate descriptive captions for images. By leveraging a pre-labeled image dataset, this project aims to improve accessibility for visually impaired users and enhance applications in HEALTHCARE - by Generating descriptions of X-ray or CT scan images to assist doctors.. The model will learn to correlate image features with relevant textual descriptions, ensuring accuracy and context relevance."

Question 2. Give the following for your data set,

1. **Collector(s):** The name(s) of the individual(s) or organization responsible for creating the dataset.

University of Illinois URBANA – CAMPAIGN

The Grainger College of Engineering
School of Computing and Data Science

2. **Year:** The year when the dataset was first published or made available.

(2013)

3. **Title of Dataset:** The title of the dataset in italics or as required by the citation style.

Flickr 8k Images

4. **Version Number (if any):** The version or edition of the dataset, if applicable.

5. **Publisher:** The entity that published or distributed the dataset.

M. Hodosh, P. Young and J. Hockenmaier

6. **DOI or URL:** The Digital Object Identifier (DOI) or the direct URL to access the dataset.

Flickr 8k Images

Here https://linkprotect.cudasvc.com/url?a=http%3a%2f%2fnlp.cs.illinois.edu%2fHockenmaierGroup%2fFraming_Image_Description%2fFlickr8k_Dataset.zip&c=E,1,o72OaLlhX03naqrk7CkCqQ-bM3mMDFjxF41YBYf-hiGIVqTrTfX-eb3ymDmEv_p4dNaFYtDSVmWS6mre7XQHbO5cblw4FtDyGdO90nUkP95e6tkQEFzK9QQ.&typo=1

Flickr 8k Text

Data: https://linkprotect.cudasvc.com/url?a=http%3a%2f%2fnlp.cs.illinois.edu%2fHockenmaierGroup%2fFraming_Image_Description%2fFlickr8k_text.zip&c=E,1.jp2uhdpP6QhWHQnRml2r7TKyLbWaZrj3l_HiObqRJZ1XlPe8C_ebTkAT3rHAXOcErKrc6lU4w2ANz1U_zcwgXm9PMHWle0b_9zW-bUxVt1hkrNjg5i70Bw8,&typo=1

7. **Study/Paper/Reason:** The original reason the data was collected.

I am still working on datasets, I got some data sets for image captioning but I am in a thought to combine image captioning datasets to that of healthcare images datasets so as to advance my project.

Once I figure out what datasets to combine according to my research I will push that to my github so that you can have a look.

Question 3. List the language and the libraries you'll use. E. g., Python 3.13.1, Numpy, Matplotlib etc.

Language: Python

TensorFlow, PyTorch, Numpy, Matplotlib, OpenCV, Natural Language Toolkit (NLTK), Transformers (Hugging Face).

Question 4. Describe what code will you be writing yourself?

Custom preprocessing pipeline for image resizing and augmentation.

Data loader scripts to map images to captions.

Implementation of a deep learning model combining Convolutional Neural Networks (CNNs) for feature extraction and Recurrent Neural Networks (RNNs) for sequence generation.

Question 5. What do you think is the best choice of model(s) for your project? Why?

"I will use an encoder-decoder architecture combining CNNs (like ResNet) to encode image features and RNNs (LSTM or GRU) for generating captions. Alternatively, I am considering transformer-based models like Vision Transformers (ViT) or CLIP for state-of-the-art results. These models provide robust feature extraction and contextual text generation capabilities."

Question 6. What are the hyper-parameters of this model and how will you pick the optimal ones?

The hyperparameters will be based on models I will be using. Like learning rate, batch size, number of layers if RNN used.

Question 7. How will you evaluate the performance of the model?

I will be using metrics like precision, recall and f-1 score and combine them accordingly with bleu or any other metrics via research while working on this project.

Question 8. You'll need to learn Git and GitHub to some extent. Here are two resources:

1. [Git Tutorial For Dummies \(Youtube Video\)](#)

2. [Git de Facto Reference](#)

Start a Git repository and setup a GitHub origin for it. It should contain and be named the following:

```
lastname-bid-semester-year      # E. g.,. doe-b01234567-fall-1970
├── data                         # Your data(s). See README.md if it is more than 2GiB.
│   └── data.csv
├── deliverables                 # This directory should contain all your documents.
│   └── presentation            # Anything related to your final presentation.
│       └──                    # Your final presentation PDF.
presntation.pdf
│   └── final_idea.pdf          # Your final project idea.
│       └── three_ideas.pdf     # Your original three ideas document.
├── src                          # All code must be contained here.
│   └── program.py
└── README.md                   # Explain how can we run your code on our machines. Or,
                                obtain a copy of your data if not in your repository.
```

Your progress throughout the semester will be judged based on the work done in this repository per commit. To help with Git or GitHub, send the professor an email.

GitHub link - <https://github.com/Sai-Chandra-Sriramula/Sriramula-B00116842-2025>

Submission Instructions

1. Submit a PDF answering all the above questions. Make sure to include the GitHub link for the question [8](#).
2. Also leave the GitHub link from question [8](#) in the “Comments” section of the D2L Dropbox.

Oklahoma City University, Petree College of Arts & Sciences, Computer Science