



## Introduction

**Our project was to develop a system capable of:**

1. Using pre-trained image classification model & LSTM to tag images
2. Applying sentence similarity algorithms to compare and filter an image database based on tags
3. Making the database search system intuitive and accessible to users



## Motivation

### Captioning for the visually impaired:

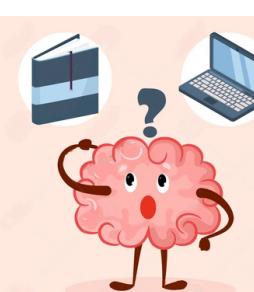
- Auto-captioning systems are amazing for building a more inclusive society.

### Image search optimization:

- Transitioning from a image-based feature space to a text-based one
- Search engines better understand image content and provide more relevant search results.

### Database generation:

- Build a smaller subset of image from larger set using user keywords

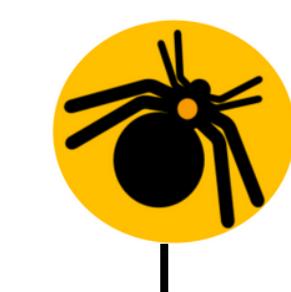


## Methodology

- **Feature Extraction:** A CNN extracts visual features from the input image, resulting in a high-dimensional vector.
- **Dimensionality Reduction:** A fully connected layer reduces the dimensionality of the feature vector to obtain an image embedding.
- **Caption Generation:** An RNN, such as an LSTM, generates a descriptive text sequence from the image embedding.
- **Word-by-Word Generation:** The LSTM generates captions a word at a time, based on the image embedding and previously generated words.
- **Training:** The entire encoder-decoder framework is trained end-to-end on a large dataset of image-caption pairs using backpropagation and gradient descent.
- **Evaluation:** The quality of the generated captions is evaluated using metrics such as BLEU score.
- **Fine-Tuning:** The pre-trained encoder-decoder framework can be fine-tuned on a specific task using a smaller dataset of images and captions.



## Data Pipeline



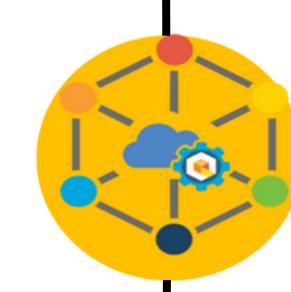
### Data Collection:

- Read in the unlabelled image dataset.
- Read in the image captions from the Flickr30k dataset.



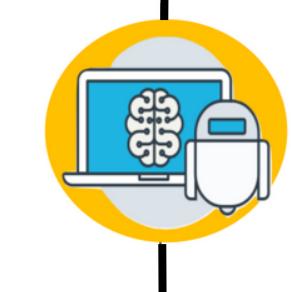
### Data Cleaning:

- Preprocess images by resizing and normalizing them and extract image features using a pre-trained CNN such as DenseNet 201.
- Tokenize and preprocess user-defined phrases and image captions by converting to lowercase, removing special characters, numbers, and punctuation marks, and adding start and end tags. Extract relevant tags using natural language processing pipeline.
- Combine the extracted tags from user-defined phrases and image captions to create a vocabulary of relevant tags.



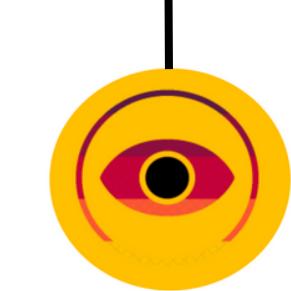
### Data Integration:

- Encode the relevant tags and the image features as input to multiple neural networks with different feature spaces.



### Data Science:

- Train the neural networks to filter the unlabeled image dataset based on the user-defined phrase.
- Evaluate the efficacy of the different neural networks and select the best-performing one(s).
- Apply the selected neural network(s) to filter the unlabeled image dataset and generate the final user-requested dataset of images that match the user-defined phrase.



### User Experience:

- Built a frontend using the Streamlit library that takes text or images as input & output images with similar visual characteristics.

## Future Scope

We will evaluate the effectiveness of other trained models using the Fuzzy ratio method. This method will measure the similarity between the predicted and actual outputs, allowing us to compare the accuracy of different models.



## Image Tagging

dog is running through the water

rock climber is climbing rock

bike biker in the woods

boy in blue shirt is playing in the air

man in red shirt is standing in the grass

man in red shirt is running in the grass

girl in blue shirt is playing in the air

is playing soccer

man in black shirt and black shirt is standing on the street

woman in blue shirt and black shirt and woman in black dress

man in red shirt is walking down the street

man in red shirt is sitting on the street

dog is running in the grass

rock climber is climbing rock

man in red shirt is standing in the street

shirt sitting man is white wearing

people are sitting is man looking

is man green green pants

people are sitting is street

shirt boy is green green

are men boat

water people are sitting is boat

shirt man is hair hair woman

are playing men

shirt man is hair white woman

standing people are sitting building is

standing people are building

rock man is looking

hockey players hockey players

people are sitting young

water people are sitting young

people are sitting young

people are sitting young

blue shirt man is green looking

blue shirt man is running white

people are sitting building

blue shirt man is green woman

people are sitting building

people are sitting young

people are sitting building

walking people are building

people are sitting young

people are sitting young

people are sitting building

standing people are sitting man is street

standing shirt man is wearing

shirt man is looking woman

water are dogs running

shirt man is green looking

black blue shirt playing is man

black blue shirt boy playing is wearing

shirt boy is green wearing

people are sitting man is looking street

is man running dog

standing people are sitting man is looking street

people are sitting man is looking street

is man running dog

shirt man is looking woman

man in red shirt is standing in the grass

man in red shirt is standing in the grass

man in red shirt is standing in the grass

man in red shirt is standing in the grass