**Project Title: Image Caption Generation Using Deep Learning**

**Abstract**: The AI-based automatic image captioning project using deep learning will develop systems that provide meaningful captions for images. Computer vision and natural language processing will be employed through VGG16-based feature extraction for images and sequence generation using LSTMs. It is trained using the Flickr8K dataset: 8,000 images with textual descriptions. It will analyse the important visual features of the input images and express these features in coherent text captions. The evaluation of the project will consider both BLEU and ROUGE scores that measure caption correctness and contextual relevance. Results would provide a benefit to AI accessibility, indexing of content, and intelligent image search. Improvements in the future include better contextual accuracy via attention mechanisms and bigger datasets.

**Project Objectives:** This project aims at automating image captioning to give a textual description of images that looks very human-like. The aim of this system is to improve the access for visually impaired users, to enhance image retrieval, and to provide meaningful context to images in digital platforms. Also, the goal is to improve the balance of computational efficiency and caption accuracy by refining the architecture and the training strategies of the model.

**Methodology:**

Dataset Selection: The Flickr8K dataset is used for model training and evaluation.

Feature Extraction: VGG16 extracts high-level visual features from input images.

Caption Generation: An LSTM-based decoder processes extracted features to generate captions.

Evaluation: BLEU and ROUGE scores assess caption accuracy and semantic relevance.

**Key Findings:**

The model successfully generates relevant captions with moderate linguistic coherence.

BLEU and ROUGE scores indicate semantic alignment but highlight challenges in capturing complex linguistic structures.

Future improvements include integrating attention mechanisms and expanding dataset diversity.

**Stepwise Solution Approach:**

Step 1: Preprocess the dataset, including text normalization and image resizing.

Step 2: Train VGG16 for feature extraction and LSTM for sequence modelling.

Step 3: Optimize model parameters and evaluate performance using BLEU and ROUGE metrics.

Step 4: Refine model architecture for improved caption quality and coherence.

**Reference:**[1] Shan-E-Fatima et al., "Image Caption Generation Using Deep Learning Algorithm," Educational Administration: Theory and Practice, 2024.

**Team:**

E22CSEU1183 Sanchit Mishra

E22CSEU1181 Anurag Munjal