

# EDS 6352 - NATURAL LANGUAGE PROCESSING

## FINAL PROJECT PROPOSAL

**Title:** English QA: Transformer-Based Extractive Question Answering using SQuAD v2

**Team Members:** Group - 9

- Adi Karthikeya Sunkara Balasubramanyam - 2417324
- Amrutha Vaishnavi Alla - 2378344
- Gayatri Chekuri - 2404773

### Project Description

Question Answering (QA) is a core task in Natural Language Processing that tests a model's ability to comprehend text and locate precise answers to questions. This project will develop an extractive QA system in English using the SQuAD v2 dataset, which contains both answerable and unanswerable questions. We will fine-tune and compare modern transformer architectures—BERT, DeBERTa, and RoBERTa—and optionally evaluate a Flan-T5 model to explore a generative approach. Our goal is to analyze model performance in identifying accurate answer spans and handling unanswerable cases.

**Data Source:** We will be using the dataset Stanford Question Answering Dataset(SQuAD v2) which has over 140000 QA pairs, so plan to use a subset of 10,000 QA pairs for efficient training and evaluation within Colab GPU limits. Dataset link: <https://rajpurkar.github.io/SQuAD-explorer/>

### Approach

- Preprocess the dataset into structured (question, context, answer) triples. Each model's native tokenizer will then be applied during its respective fine-tuning stage.
- Fine-tune each model on the subset and evaluate using Exact Match (EM), F1-score, and BERTScore to assess both lexical and semantic alignment between predicted and reference answers.
- Adi Karthikeya S B will fine-tune the BERT-base model, handle data preparation, and record baseline results.
- Amrutha Vaishnavi Alla will fine-tune the DeBERTa-base model, optimize hyperparameters, and compare against the BERT baseline.
- Gayatri Chekuri will fine-tune the RoBERTa-base model, visualize metrics, and assist with final documentation and demo.
- Optionally, we will run a Flan-T5-base model for a generative comparison to observe how sequence-to-sequence models handle QA.
- As a team, we will analyze strengths, weaknesses, and error cases to propose future improvements.

### Computational Resources

- Development and training will be performed in Google Colab Pro with GPU acceleration.
- All code, training, and evaluation will be managed in Jupyter Notebooks for reproducibility.