

Final Project Proposal - Group 1  
Carrie Magee, Paul Kelly, Jack McMorrow, Akshay Verma

After you have selected a topic, a model, and a data set, submit a proposal of what you plan to do for the project. The proposal should be a few hundred words, and should address the following items.

- What problem did you select and why did you select it?
  - While computers are renowned for their proficiency in processing and communicating via numbers, real-world challenges often involve more complex components, such as text and context, that surpass their innate capabilities. Our goal is to develop an LLM model that can solve mathematical word problems, bridging the gap between computational mathematics and abstract reasoning.
- What database/dataset will you use?
  - The dataset we will be using is [SVAMP](#) (Simple Variation on Arithmetic Math word Problems). The dataset was created after models were found to rely on shallow heuristics in benchmark word math problem datasets like [ASDiv-A](#) and [MAWPS](#). The SVAMP dataset contains 1000 word problems, we will be using this data set for evaluation. The ASDiv-A and MAWPS dataset contains 1921 and 1217 word problems respectively.
- What NLP methods will you pick from the concept list? Will it be a classical model or will you have to customize it?
  - We plan to employ a combination of fundamental NLP methods and a customized model. Our preprocessing pipeline will include essential techniques such as tokenization, lowercasing, and special character removal to prepare the text data for model input. The core model will use the RNN Seq2Seq Basic Encoder-Decoder with Attention Network.
- What packages are you planning to use? Why?
  - We plan on using Tensorflow, Keras for model building. Gensim for pre-processing. We will also use Numpy, Matplotlib, Pandas, Scikit-learn for other tasks like visualization or sampling.
- What NLP tasks will you work on?

- We will focus on data preprocessing tasks such as tokenization, stopword removal, and embedding techniques. Additionally, we will engage in Sequence-to-Sequence (Seq2Seq) Generation, a specific natural language processing task, to develop our predictive models.
- How will you judge the performance of the model? What metrics will you use?
  - We will be using Accuracy, BLEU Score (Bilingual Evaluation Understudy Score) for the evaluation. When using LLM for machine translation or text generation tasks, the BLEU score can be used to measure the similarity between the model's output and our reference text problems.
- Provide a rough schedule for completing the project.
  - Nov 5 - finish project proposal
  - Nov 20 - Have models done & write a rough draft of the final report
  - Nov 27 - Make edits to final report and finish deliverables
  - Dec 4 - Finalized group report, deliverables, write individual reports
  - Dec 11 - final project due