

**OPEN DOMAIN QUESTION ANSWERING****Project Idea:**

Machine comprehension is a challenging NLP task that focuses on understanding written language. Question Answering is a significant part of machine comprehension and the aim of this project is to build an open domain question answering system using deep learning techniques. The 'bot-like' system will be designed to answer questions on a wide range of general topics by understanding the context of the question from vast amounts of data which is a challenging task.

The two primary tasks are to determine when no answer is available (key problem posed by SQuAD 2.0) and to correctly return the span of text which answers the question when possible.

**Dataset:**

The Stanford Question Answering Dataset 2.0 ([SQuAD](#)) is an extension of the original SQuAD 1.0 dataset, which includes questions that do not have an answer within the given text which makes it a challenging task. SQuAD is a large-scale dataset of question-answer pairs that are based on Wikipedia articles. It contains over 100,000 questions and answers, and is split into a training set, a development set, and a test set.

Additionally if ensemble methods are used to boost performance we can consider using the NarrativeQA by Google DeepMind or the [Quasar](#) datasets.

**The Approach:**

Unlike some other machine learning models that may be more general-purpose, 'Bidirectional Attention Flow' [BiDAF](#) has been specifically designed to tackle the challenges of question-answering. While this is a strong model, many research papers on this topic suggest that it does not use pretrained contextual embeddings (non-PCE model) which is supposed to yield better results. Therefore this makes BiDAF a promising baseline which can be compared to newer state of the art models like BERT.

Instead of using fixed embeddings for words, newer models such as BERT leverage the context that the word appears in. Additionally BERT has the ability to handle out-of-vocabulary words which makes it an extremely useful solution specifically for the SQuAD dataset because it tests the ability to abstain when presented with an unanswerable question.

By fine tuning BERT I believe we can obtain significantly good results than conventional models.

The initial proposed methods of the evaluation of this model will most likely be F1 scores and Exact match scores based on SQuAD guidelines.

**Softwares/ Libraries used:**

Frameworks such as Pytorch, Tensorflow for deep learning

**References:**

1. Reading Wikipedia to Answer Open-Domain Questions  
Danqi Chen, Adam Fisch, Jason Weston, Antoine Bordes <https://arxiv.org/abs/1704.00051>
2. Y. Yang, "BiEAF: An Bidirectional Enhanced Attention Flow Model for Question Answering Task," *2021 2nd International Conference on Information Science and Education (ICISE-IE)*, Chongqing, China, 2021, pp. 344-348, doi: 10.1109/ICISE-IE53922.2021.00086.
3. Know What You Don't Know: Unanswerable Questions for SQuAD  
Pranav Rajpurkar, Robin Jia, Percy Liang <https://arxiv.org/abs/1806.03822>

**Timeline:**

- Initial baseline model implementation by 05/22/23
- Future Models Evaluation and Optimization by 06/01/23
- Presentation ready by 06/08/23