# Reflection of Research Experience

My URECA research topic, Deep Learning with Question Answering (QA) Systems, resides in the field of Natural Language Processing (NLP). Initially, I decided to embark on this journey because I wanted to broaden my skills and understanding in the field of Artificial Intelligence (AI). NLP, a growing sub-topic in AI, piqued my interest. I would describe my experience as an arduous, confusing, yet fruitful one. Though there were many things I had learnt, there were things I think could have went better.

The initial phases of URECA were fraught with confusion. As I had little prior knowledge of NLP and QA systems, I spent large portions of my times scouring through an ocean of information on the web to learn about this mysterious topic of QA systems - blogs, books, videos etc. After a long and difficult time of research, I found a few very good resources about QA systems. One of which was a well-written e-book by Chris McCormick called “The Inner Workings of BERT”. This e-book set a significant milestone in my URECA research as it gave me a solid understanding of the BERT architecture. This meant that I could start reading related research papers confidently without the terrible feeling of confusion I had before. The field of NLP is extremely complex; there were a lot of dependencies in topics. For instance, to learn about the BERT architecture, one had to know about the Transformers architecture, which is a rather unrelated topic to QA system on its own. This made it extremely difficult for a fresh URECA student to learn the initial ropes of QA systems. If I could go back in time, I would give my past self this exact e-book to read as it would have saved so much time!

Another significant milestone was when I discovered Haystack an open-source framework for building search systems. Learning how to use this software enabled me to build an end-to-end QA system on my local computer. First, text documents could be indexed into ElasticSearch using Haystack. After which, I could implement the ‘retriever-reader’ architecture. There were other functionalities like text processing, model training and evaluation that allowed me to learn the inner workings of BERT more deeply. Though the process of learning this software was by no means easy for a beginner either, implementing a QA system gave me confidence and inspired me to learn more.

As the poster submission neared, I was tasked by my mentors to come up with a problem statement that could be solved using QA systems. I explored with many ideas and after much deliberation, came to the idea of building a question answering system that answers about covid-19 rules and regulation in Singapore – a Covid-19 QA system. As covid-19 rules and regulation in Singapore are rather extensive and they change from time to time, the objective was to make it easy for public to answer their queries about this information. Thus, I began scraping and collecting these regulatory documents or other relevant sources of documents. These were then indexed into ElasticSearch as the datastore to complete my covid-19 QA system.

Another noteworthy experience was learning how to use HuggingFace - an open-source provider of NLP. Though I was introduced to this software very early on, the lack of proper understanding and knowledge of Transformers proved to be a huge barrier. It was not until the second half of the URECA stage that I had enough understanding and found some good tutorials to learn this crucial NLP software. Compared to Haystack, HuggingFace had more essential functionalities and hyperparameters for me to explore with. This enabled greater customization and capacity for experimentation for my final research paper.

The URECA research program gave me direction to learn about the field of NLP. I have learnt many new skills and knowledge like reading research papers, web scraping, HuggingFace, BERT and Transfomers that I never had learnt otherwise. I can confidently say that the journey was a fruitful one. However, this is by no means the end. I still have a lot to learn, and I will continue to build on the skills and knowledge I have gained from URECA.