# **Project Proposal**

### UIUC: Text Information Systems

Group Name: Team Coco

Theme: Free Topics

Specific Topic: AllenNLP Chatbot

Captain/Project Leader (and sole team member): Joel Haas, [joelah2@illinois.edu](mailto:joelah2@illinois.edu)

**Proposal Task**

**Configure a chatbot using the Allen Institute for Artificial Intelligence open-source library AllenNLP**

## Description

AllenNLP is an open-source library developed by the Allen Institute for Artificial Intelligence. The library supports developers wanting to build high quality deep learning NLP models. The Allen Institute’s goal was to simplify the deep learning NLP model building process. They do this by providing an abstraction layer for common components and models. This abstraction layer enables data scientists to easily configure, run, and manage NLP experiments. The abstraction layer is also intended to remove the developer from the underlying deep learning complexities, reduce the amount of code that needs to be written, and at the same time, increase re-use and flexibility of configured NLP pipelines across different use cases.

The libraries are built on top of PyTorch, and AllenNLP provides APIs for common NLP tasks, so the developer doesn't need to write code for them. Examples include tokenizing the text, converting each text ID into a vector, and then creating a single vector for each input text, where the output of each is a tensor. These pre-processing steps are encapsulated in a Model constructor and AllenNLP handles the various data mappings, persistence, and retrieval. Users can adjust these Model constructor parameters as needed.

**In this project, I want to explore the AllenNLP library and how to configure a deep learning NLP pipeline specifically focused returning answers to user questions.**

## Additional Information

Interest:The Allen Institute for AI developed this library to ease the pain of configuring deep learning NLP pipelines so developers can focus on tackling hard AI problems. I want to explore the library and how the abstraction layer works. If it really is as they describe, it could be ground-breaking for NLP developers.

## Programming Language: Python

Tools: AllenNLP library

Systems: Personal laptop

Datasets: Will need to find one to use

Expected Outcome: Deep learning NLP pipeline for a Chatbot

Evaluation Methodology: Manual inspection + comparing against existing AllenNLP chatbots

Expected Workload by Task:

5 hours - AllenNLP research

5 hours - Configure and test simple AllenNLP pipeline

2 hours - Find datasets for training/testing

15 hours - Configure and debug deep learning NLP pipeline for a chatbot

5 hours - Evaluate performance and iterate

5 hours - Documentation and reporting requirements