# CSCI-GA 2590 Project Proposal

Team Member: Hongyi Zheng, Jiarao Liu, Yijie Wang, Zhitong Guo, Yi Wei

## Overview

What problem are you going to work on?

Currently, large language models (LLMs) have achieved impressive performance in natural language tasks such as text summarization, question answering, and logical reasoning in general, but when it comes to domain-specific tasks which require intensive domain knowledge and commonsense reasoning, even the most sophisticated LLMs may fail. In order to tackle this issue, we plan to build a chatbot restricted to a specific domain and infuse necessary domain knowledge into the conventional LLMs.

What are the challenges?

In order to generate accurate answers, language models need to incorporate factual domain knowledge in the generative process. One of the main challenges is designing a new paradigm to infuse the domain knowledge with the pre-trained model. In addition, working with different types of datasets could be difficult as they are from other domains.

What's your solution?

We plan to use Graph Reasoning Enhanced Language Models for Question Answering (GreaseLM), which is a model that reasons over the integration of knowledge graph (KG) representation and language models. The model initially learns representations of input tokens, and then mixes the graph representations of KG with textual representations to robustly represent relationships between different concepts.

## Project Plan

What do you plan to do (experiments, data, model)?

We will investigate common knowledge-infusion paradigms and propose a new fusion method that builds on top of them. (pre-fusion/post-fusion/hybrid-fusion). The language models we plan to use to generate input embeddings are the most representative state-of-the-art language models (e.g. GPT-3), and the model for knowledge distillation and representation is rather diverse (e.g. GreaseLM, ERNIE).

For the dataset, we will mainly focus on domain-specific datasets such as *SOCIALIQA* for social commonsense, *PIQA* for physical commonsense, as well as other domain-specific question-answering datasets.

How do you evaluate success?

We will primarily evaluate the performance of our approach with the standard evaluation metrics such as accuracy/F1-score for binary/multi-class classification tasks (e.g. math problem answering), GLEU score for text generation tasks (e.g. text summarization). The main goal of our project is to investigate novel approaches to enhance the system's performance over domain-specific datasets and surpass the benchmark of state-of-the-art solutions.

**References**


GreaseLM: Graph REASoning Enhanced Language Models for Question Answering

A Survey of Knowledge-Intensive NLP with Pre-Trained Language Models

AdaptSum: Towards Low-Resource Domain Adaptation for Abstractive Summarization

Efficient Transformers: A Survey

Everything Has a Cause: Leveraging Causal Inference in Legal Text Analysis

Domain Knowledge Transferring for Pre-trained Language Model via Calibrated Activation Boundary Distillation

Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model