Question-Answering System for the Insurance Industry: Project Report

Introduction

This project aimed to develop a question-answering system for the insurance industry using LangChain and the Falcon-7b-Instruct large language model (LLM) from the Technology Innovation Institute. The system was designed to provide accurate and context-aware answers to user queries based on a dataset of published insurance papers. This methodology is adaptable and applicable to various text corpora.

Methodology

Data Collection:

A diverse dataset of published insurance papers was collected and preprocessed to ensure consistency and relevance. The dataset served as the foundation for training and evaluating the question-answering system.

Model Selection:

Falcon-7b-Instruct was chosen as the primary LLM due to its capabilities and suitability for natural language understanding tasks within the insurance domain. The model's architecture and specifications influenced the subsequent fine-tuning process.

Integration with LangChain:

LangChain was seamlessly integrated into the system to facilitate user interaction with the Falcon-7b-Instruct model. The logic for processing user queries and delivering relevant answers was implemented using LangChain's capabilities.

User Interface:

A user-friendly interface was designed to enable easy interaction with the question-answering system. Features such as natural language input and intuitive navigation were incorporated to enhance the user experience.

Challenges

Computational Resources:

The use of Falcon-7b-Instruct, a large language model, required significant computational resources. Optimization techniques were employed to manage resource constraints and ensure efficient model training.

Results

Model Performance:

The fine-tuned Falcon-7b-Instruct model demonstrated promising results, achieving high accuracy in providing relevant answers to insurance-related queries. Comparative evaluation against other models or baselines confirmed Falcon-7b-Instruct's effectiveness in the insurance domain.