## Abstract

(Contents of the outline /abstract is to be given here)

##### Abstract (in about 500 words)

The objective of this dissertation is to enhance the Annual Compliance Confirmation (ACC) process using advanced Artificial Intelligence (AI) techniques such as Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), and Predictive Analytics. The focus is on automating the process of answering compliance-related questions using a Predictive analysis, identifying the likelihood of false positive responses in submitted answers, thereby minimizing risk and improving decision-making accuracy.

The project is structured around two core problem statements. The first involves developing a conversational AI agent that utilizes Retrieval-Augmented Generation (RAG) techniques combined with Large Language Models (LLMs) to respond to compliance queries using information extracted from a knowledge base maintained in Azure DevOps Wiki. This includes ingesting semi-structured knowledge, indexing, and fine-tuning the LLM to ensure relevance, accuracy, and traceability of responses.

The second problem statement addresses risk prediction by leveraging machine learning techniques to identify false positive responses to ACC questions. A dataset comprising 15,000 classified false positive cases is available for analysis. Each response has been categorized into one of three types: Exception to Policy, False Positive, and Inquiry Addressed. To maintain employee privacy, identifiers are hashed.

Initial stages of the project will involve extensive data cleaning, normalization, and exploratory data analysis to ensure data quality and usability for model training. A comparative study of different classification models—Random Forest, Decision Tree, Logistic Regression, Gradient Boosting Machine (GBM), and Support Vector Machine (SVM)—will be conducted to determine the most effective classifier in predicting false positives.

This dissertation explores the intersection of conversational AI and risk prediction in compliance systems. It aims to streamline ACC responses, reduce human workload, and proactively identify compliance risks. The resulting models and systems will support compliance officers by flagging potentially unreliable answers and automating repetitive question resolution, contributing to a more transparent and robust governance framework.

**Key Words**:  
Annual Compliance Confirmation, NLP, RAG, Predictive Analytics, Machine Learning, False Positives, Risk Assessment, Azure DevOps, Data Cleaning, Classification Models