# Proposal: Integrating a Neo4j Knowledge Graph with the Golf Rules RAG System

## Introduction and Background

The current Golf Rules Q&A system leverages Retrieval-Augmented Generation (RAG) to answer questions. It uses an LLM with a vector-based retriever over two data sources: (1) the official Rules of Golf (embedded from a PDF rulebook) and (2) a collection of thousands of email Q&As (embedded question-answer pairs)...

## Rationale for Knowledge Graph Augmentation (Not Replacement)

Introducing a knowledge graph is motivated by the limitations observed in the current RAG approach and the potential gains from structured knowledge...

## Hybrid System Architecture Overview

In the proposed solution, the overall architecture becomes a hybrid RAG + KG system. The high-level flow is as follows: 1. User Query Ingestion...

## Knowledge Graph Schema Design for the Rulebook

Structuring the entire rulebook into a graph requires defining what the nodes and edges represent...

## Building the Knowledge Graph: Extraction and Ingestion

Constructing the knowledge graph from the unstructured rulebook is a critical step. We plan to use a combination of LLM-based parsing and symbolic rule mining...

## Graph-Enhanced Retrieval for Improved Context Selection

With the knowledge graph in place, the retrieval process for answering questions will be augmented to use it. Here we detail how graph traversal and subgraph retrieval improve the LLM’s context selection...

## Implementation Plan and Phased Rollout

Implementing this hybrid RAG+KG solution will be tackled in phases to manage risk and demonstrate value quickly. Below we outline the phases, including a rapid prototype and subsequent full-scale development...

## Value Proposition and Conclusion

In conclusion, this proposal outlines a clear path to integrate a Neo4j knowledge graph with the existing RAG architecture to create a more powerful Golf Rules question-answering system...