How to Develop Explainable and Domain-Adaptive RAG Systems for Cybersecurity Intelligence

Muhammad Asif Nawaz
Software Industry

Abstract

Retrieval-Augmented Generation (RAG) systems have emerged as an effective remedy to the limitations of Large Language Models (LLMs), including their parametric memory limitations and a tendency toward hallucinations. By including non-parametric knowledge retrieval (querying external databases), RAG equips models with an enhanced external memory. However, current RAG systems encounter two significant challenges that limit their deployment in critical, niche domains like cybersecurity: 1. Domain Adaptation failure during rapid knowledge changes. 2. Lack of Explainability (how retrieved documents

influence the generated output? Because of this, security analysts may find it difficult to trust the model's decisions.

This research aims to develop an RAG framework specifically for cybersecurity that focuses on two areas: first, developing dynamically adaptive retrieval mechanisms that prioritize the most recent threat intelligence and incorporate a temporal decay factor; second, applying post-hoc attribution techniques to produce explanations that are auditable and comprehensible to humans. In this research, I will utilize publicly available datasets, such as CVE/NVD feeds and OSINT reports.

The cybersecurity prospect is highly adversarial and regularly updated. While new software flaws (common flaws and exposures) are discovered and recorded every day, attackers regularly upgrade their tactics, methods, and procedures (TTPs). Consequently, security analysts must perpetually stay up to date with new developments. This continual

information is referred to as “Rapid knowledge change”, and it represents one of the most significant problems in cybersecurity operations.

Index Terms

Large Language Models (LLMs), Retrieval Augmented Generation (RAG), Cybersecurity, Domain Specific, Knowledge Source, Retrieval Augmented, Zero-Day Vulnerability