COMPUTER AND CONTROL ENGINEERING

RMA - Integration and optimization of retrieval-augmented generation (RAG) in large language models (LLMS) for the insurance industry

Funded By	SOCIETA' REALE MUTUA ASSICURAZIONI [P.iva/CF:00875360018]

Supervisor	MORRA LIA - lia.morra@polito.it

Contact	LAMBERTI FABRIZIO - fabrizio.lamberti@polito.it MORRA LIA - lia.morra@polito.it

Context of the research activity	Large Language Models (LLMs) face challenges like hallucination, outdated knowledge, and opaque reasoning. Knowledge Graphs (KGs) store rich factual knowledge in a structured way, with the potential to improve LLMs' robustness and interpretability, but are labour intensive to build and maintain. This research proposal aims to integrate KGs and LLMs to develop robust Retrieval-Augmented Generation (RAG) applications for the insurance sector. It considers aspects such as the combination of LLMs with graphs, the optimization of relevant content retrieval, the verification of factual consistency, the evaluation of RAG's effectiveness in different contexts.

Objectives	Large Language Models (LLMs) showcase impressive linguistic capabilities but produce hallucination, may incorporate outdated knowledge, and their predictions are non-transparent as the underlying "reasoning" process is untraceable. Retrieval-Augmented Generation (RAG) has thus emerged as a promising solution by incorporating knowledge from external databases, including structured (knowledge graphs) and unstructured (textual sources). RAG has been shown to increase the accuracy and credibility of the generation, particularly for knowledge-intensive tasks, allowing for continuous knowledge updates and integration of domain-specific information. In particular, Knowledge Graphs (KGs) are structured representations that explicitly store rich factual knowledge. KGs can enhance LLMs by providing external knowledge for inference and interpretability. Meanwhile, KGs are difficult to construct and maintain over time; therefore, there is a need for automated methods in KGs to generate new facts and represent previously unseen knowledge. These limitations call for a synergistic approach in which the capabilities of both technologies are leveraged in a mutually beneficial way. Starting from a core structured representation of domain-related concepts, entities and rules, such an approach could progressively and automatically expand the KG from unstructured textual documents. The KG could then provide a generalized structured representation of existing factual knowledge, that can in turn be used to enrich generated LLM-based queries, as well as verify the factual

correctness of LLM-generated responses.

The aim of the present PhD proposal is to exploire the integration and optimization of graph-enhanced RAG in LLMs to build robust and effective applications tailored to the insurance domain. To achieve this overarching goal, the following research questions will be tackled:

- integrating information from KGs and other sources to perform RAG in LLMs for the insurance industry

- designing LLM-based systems to perform KGs update, detecting novel concepts and inconsistencies

- study to what extent RAG can improve the trusthworthiness of LLM-based applications in different contexts within the insurance industry

Skills and competencies for the development of the activity	Good knowledge of artificial intelligence, machine learning, natural language processing, and LLMs. Experience in programming and data handling. Communication and teamwork skills.