Contribution ID: **48** Type: **Poster**

# Enhancing Large Language Models with Retrieval-Augmented Generation: A Case Study on Movie Data Beyond the Training Cutoff

*Thursday 24 April 2025 10:45 (30 minutes)*

This article investigates the role of Retrieval-Augmented Generation (RAG) in enhancing Large Language Models (LLMs) with information about movies and TV series released beyond their training data. In this study, the Llama 3.2 3B LLM is leveraged and integrated with external movie-related data retrieved from the OMDb API to provide specific information about over 14000 titles released in 2024, which fall outside of the LLM's knowledge cutoff. This approach aims to improve the accuracy, reliability, and contextual relevance of LLM responses by utilizing movie metadata and precomputed embeddings for information retrieval. The incorporation of these techniques enables the system to efficiently identify plot connections, verify directors and cast members, and analyze trends in the latest movie productions. Moreover, the research examines RAG's potential in mitigating LLM hallucinations by providing reliable external knowledge and adaptive query processing. The results aim to support film critics, analysts, and movie enthusiasts by providing the latest film-related data, while also highlighting the effectiveness of RAG in fields where access to specialized, dynamic knowledge is crucial.

**Author:** MIKOŁAJCZYK, Marcel (Politechnika Warszawska, Wydział Elektryczny)

**Presenter:** MIKOŁAJCZYK, Marcel (Politechnika Warszawska, Wydział Elektryczny)

**Session Classification:** Session B (Poster)