

Leveraging Large Language Models for Enhanced Customer Support of Service Desks

Graduate



Mathias Marxer

Introduction: In this master thesis we studied the use cases of large language models (LLM) for the customer support of service desks in the IT Service Management (ITSM) domain. The primary objective was to assist support agents, improve resolution times, and enhance the scalability of service desks. A preliminary survey was conducted with ky2help® customers to identify the most relevant AI-driven use cases and assess the general acceptance of AI in ITSM.

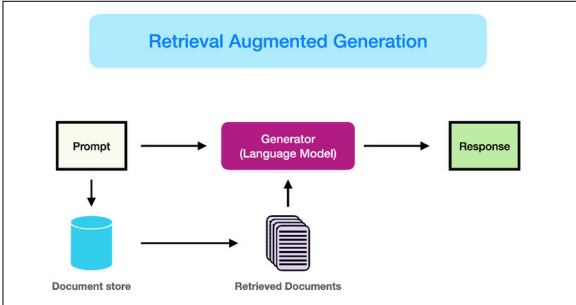
Approach: The top use case, “Support Assistant with Automated Solution Proposals”, was implemented using the LangChain LLM library across four different approaches: (1) Pretrained LLMs without internal company or customer data context; (2) A Retrieval-Augmented Generation (RAG) pipeline using Weaviate for vector storage and document embeddings; (3) A Graph RAG approach utilizing a Neo4j knowledge graph; (4) A hybrid approach combining knowledge graphs with vector embeddings.

These four implementations were evaluated using a hand curated dataset, consisting of question and answer pairs obtained from common service desk inquiries and problems from KYBERNA AG, tested with different LLMs, including Llama 3.1, Llama 3 and Aya 23. Due to data security concerns, GPT-4 and other proprietary models were not considered.

Result: The evaluation showed that the vector embeddings approach consistently outperformed others, despite the theoretical advantages of combining knowledge graphs with vector embeddings. The reasons behind this performance disparity are discussed in detail in the thesis. Despite the results, we showed that the usage of generative

AI in combination with RAG can assist support agents with automated solution proposals for inquiries that will improve ticket resolution times of service desks.

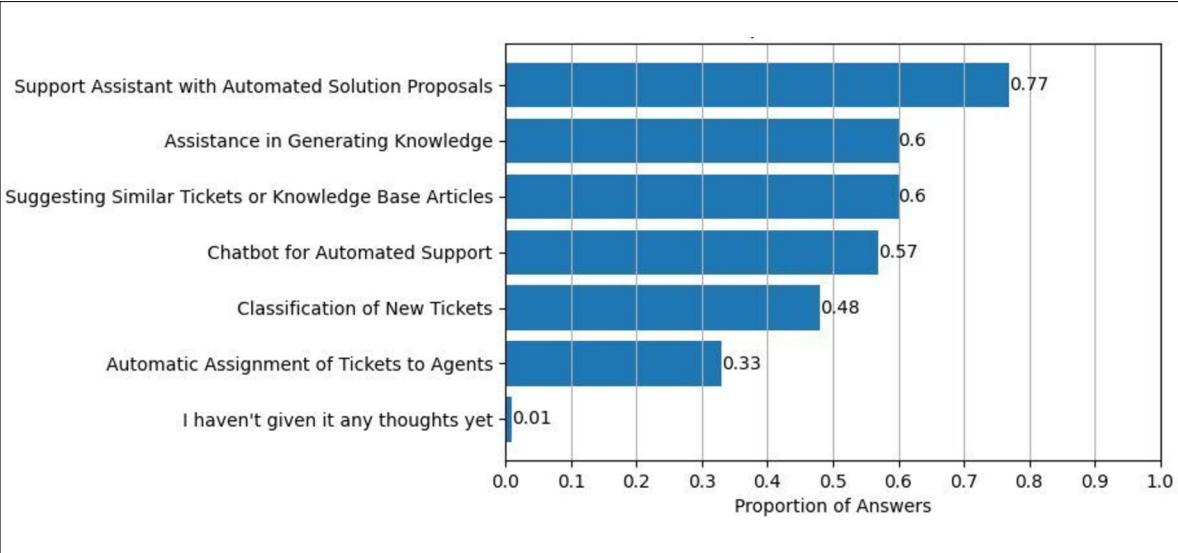
Overview of a RAG pipeline
Retrieval Augmented Generation (RAG) for LLMs, DAIR.AI



Evaluation results for llama3.1:8b, llama3:8b and Aya-23-8b as generative LLM and llama3:8b for cypher generation
Own presentment

	Cosine Similarity	BLEU	ROUGE	METEOR
Llama3.1:8b				
Pretrained LLM	0.195	0.013	0.133	0.088
Vector Embeddings	0.480	0.272	0.412	0.372
Knowledge Graph	0.239	0.017	0.155	0.099
Knowledge Graph with Vector Embeddings	0.356	0.110	0.284	0.210
Llama3:8b				
Pretrained LLM	0.198	0.013	0.136	0.092
Vector Embeddings	0.405	0.175	0.334	0.311
Knowledge Graph	0.208	0.017	0.151	0.096
Knowledge Graph with Vector Embeddings	0.282	0.060	0.224	0.185
Aya-23-8b				
Pretrained LLM	0.151	0.016	0.110	0.070
Vector Embeddings	0.370	0.165	0.314	0.219
Knowledge Graph	0.221	0.018	0.150	0.080
Knowledge Graph with Vector Embeddings	0.245	0.046	0.202	0.143

Proposed AI-driven features with percentage of votes
Own presentment



Advisor
Dr. Shao Jü Woo

Co-Examiner
Uwe Riedmann,
KYBERNA AG

Subject Area
Computer Science,
Data Science

Project Partner
KYBERNA AG, Vaduz,
Fürstentum
Liechtenstein