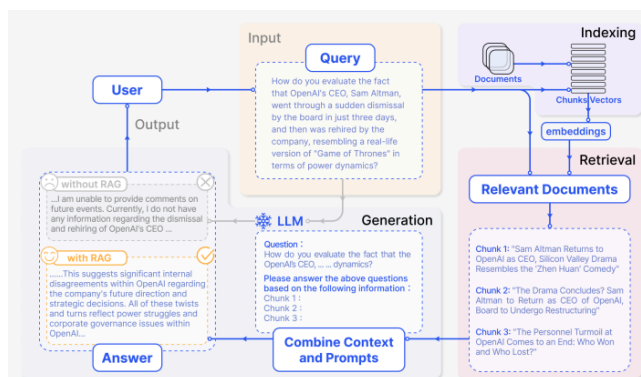


## Wie kann Retrieval-Augmented Generation für Large Language Modelle genutzt werden, um automatisch IT-Supportanfragen zu bearbeiten?

In diesem Projekt wird untersucht, wie Retrieval-Augmentes Generation (RAG) und Large Language Models (LLMs) kombiniert werden können, um die Bearbeitung von IT-Supportanfragen zu automatisieren und effizienter zu gestalten. Dabei sollen die Stärken von RAG, der effiziente Zugriff auf relevante Informationen, mit den Fähigkeiten von LLMs zur Sprachverarbeitung in einem System

vereint werden. Das Ziel ist es, für das Unternehmen badenIT eine Lösung zu entwickeln, die eingehende Supportanfragen analysiert, relevante Informationen aus bestehenden Wissensdatenbanken sowie historischen Supportanfragen extrahiert und Antworten generiert, damit der Nutzer eine Anleitung zur Problemlösung erhält oder das Model automatische die nächsten Lösungsschritte einleitet. Durch die Automatisierung dieser Prozesse soll die Bearbeitungszeit von Supportanfragen



reduziert, die Kundenzufriedenheit erhöht und der Service Desk entlastet werden. Langfristig könnte die Lösung insbesondere durch die immer mehr ansteigenden Supportanfragen einen entscheidenden Mehrwert bieten.

### Betreuer

Prof. Dr.-Ing. Janis Keuper

- [janis.keuper@hs-offenburg.de](mailto:janis.keuper@hs-offenburg.de)
- <https://www.keuper-labs.org>

### Beteiligte Institute und Firmen

Das Projekt wird in Kooperation mit der badenIT GmbH und dem Institute for Machine Learning and Analytics durchgeführt.

### Ziele des Projekts

- Wissenschaftliche Publikation der eigenen Ergebnisse
- Systematische Analyse und Evaluation der Kombination von RAG und LLM
- Anpassung und Feintuning von LLMs auf IT-spezifische
- Entwicklung eines Prototyps zur automatisierten Bearbeitung von IT-Supportanfragen

### Diese Werkzeuge/Qualifikationen werden erlernt

- Verständnis und Anwendung von Retrieval-Augmented Generation und LLMs
- Praktische Entwicklung mit Python und Frameworks wie Huggingface Transformer
- Evaluationsmethoden für Large Language Models, speziell zur Messung der Antwortqualität und Effizienzsteigerung

### Literatur + Weiterführende Informationen

- IMLA: <https://imla.hs-offenburg.de/>
- badenIT: <https://www.badenit.de>