

# Q/A Chatbot using RAG (Retrieval Augmented Generation)

**KnowRA+ by WNS Global Service**

Vidipt Vashist  
(MA22M025)



Indian Institute of Technology Madras  
Department of Mathematics  
MA5960 M.Tech Project

December 13, 2023



## ① Introduction

## ② Problem Statement

## ③ Methodology

## ④ Objective

## ⑤ References



## 1 Introduction

## 2 Problem Statement

## 3 Methodology

## 4 Objective

## 5 References



# Introduction

RAG is an AI framework for retrieving facts from an external knowledge base to ground large language models (LLMs) on the most accurate, up-to-date information and to give users insight into LLMs' generative process.

RAG specifically focuses on combining retrieval and generation AI techniques to provide context-aware responses. It excels in tasks where it needs to retrieve information from a large database and then generate coherent responses based on that retrieved data.

# Why Retrieval-augmented generation ?

- **Hallucinations:** refers to instances where the model confidently returns an incorrect or fabricated response.
- **Knowledge Cut-offs:** LLM model has a training end date, post which it is unaware of events or developments. This limitation means that the model's knowledge is frozen at the point of its last training date. For instance, a model trained up to 2022 would not know the events of 2023.

1 Introduction

2 Problem Statement

3 Methodology

4 Objective

5 References



# Problem Statement

- Organizations face a formidable challenge in unlocking the wealth of information stored within private PDF files.
- Extracting and managing this data for conversational accessibility is complex and traditional methods prove inefficient and user-unfriendly.
- The absence of efficient solutions impedes organizations from leveraging their internal knowledge effectively,
- leading to missed opportunities for enhanced customer interactions, streamlined processes, and a competitive edge in the market.

## 1 Introduction

## 2 Problem Statement

## 3 Methodology

## 4 Objective

## 5 References

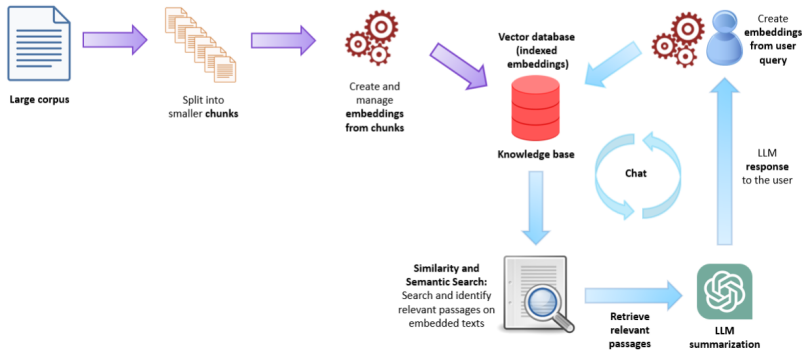




# KnowRA+ Product

KnowRA+ revolutionizes customer support by intelligently harnessing knowledge from private PDF files through an innovative chatbot. Bridging the gap in data accessibility, it ensures secure information retrieval, elevating user experience. With its robust RAG model integration and NLU capabilities.

# Methodology



## 1 Introduction

## 2 Problem Statement

## 3 Methodology

## 4 Objective

## 5 References



# Objective

- Goal of KnowRA+ is to address challenges by traditional method by providing a comprehensive solution that facilitates the creation of intelligent chatbots tailored to individual organizations' private PDF files.
- Product aims to balance security, customization, and scalability to empower businesses in leveraging their data for enhanced customer

## 1 Introduction

## 2 Problem Statement

## 3 Methodology

## 4 Objective

## 5 References



# References

- Levonian, Z., Li, C., Zhu, W., Gade, A., Henkel, O., Postle, M.E. and Xing, W., 2023. Retrieval-augmented Generation to Improve Math Question-Answering: Trade-offs Between Groundedness and Human Preference. arXiv preprint arXiv:2310.03184.
- Siriwardhana, S., Weerasekera, R., Wen, E., Kaluarachchi, T., Rana, R. and Nanayakkara, S., 2023. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. Transactions of the Association for Computational Linguistics, 11, pp.1-17.