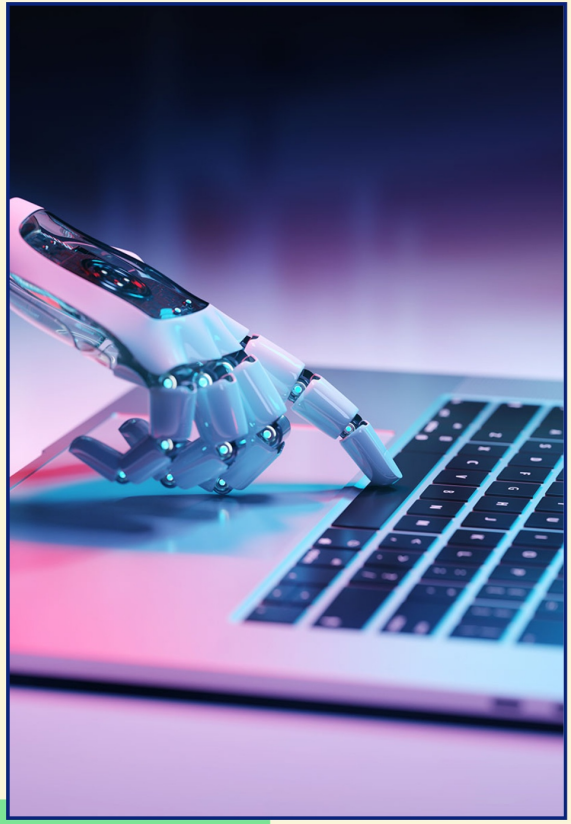
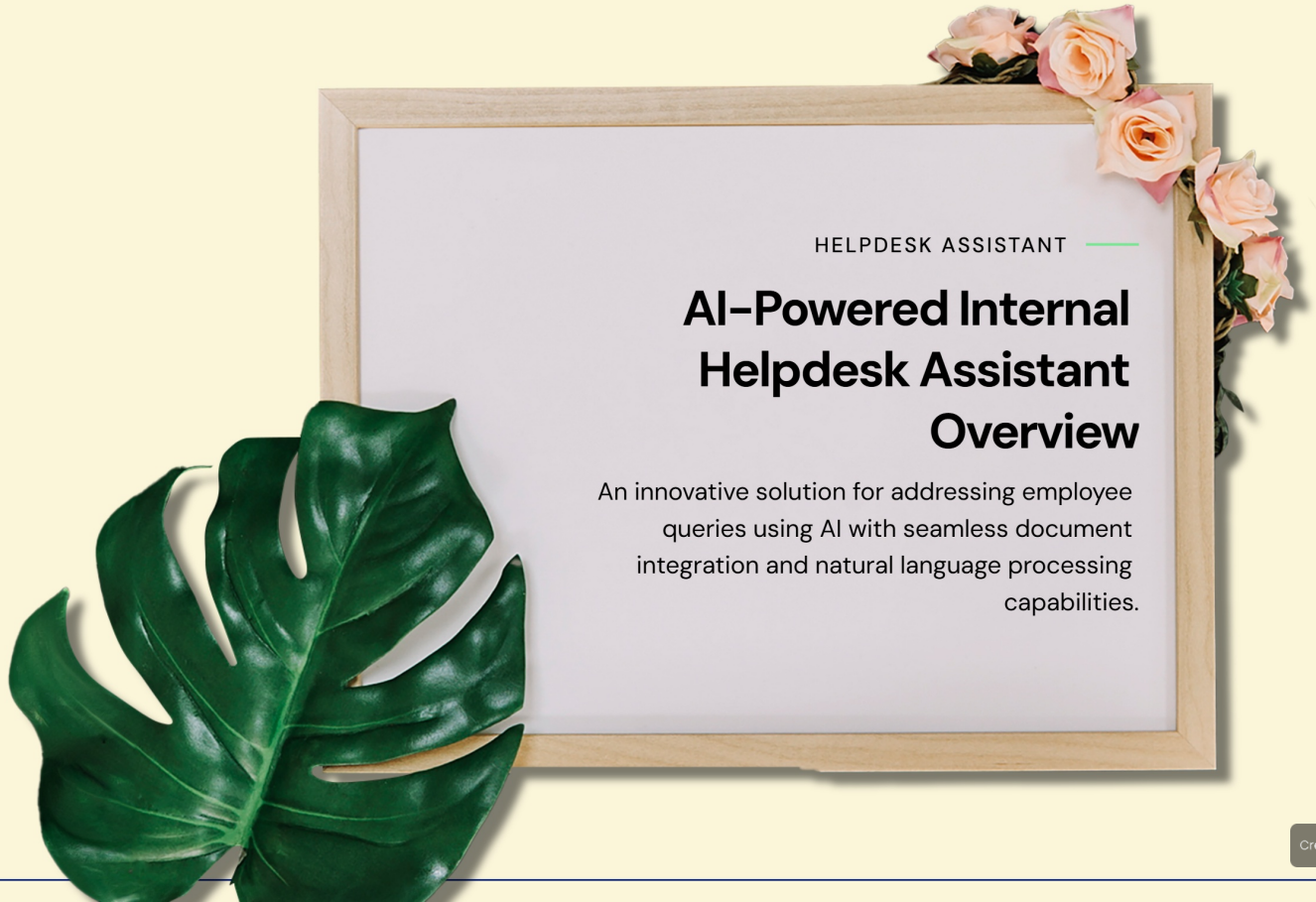


# Enhancing Employee Queries with an AI-Powered Helpdesk Assistant

This presentation introduces the **Internal Helpdesk Assistant** project, an AI-powered system designed to enhance employee query responses based on **company policies** through **document extraction** and integration of local **language models**.

**Gurupriya**
Presenter



HELPDESK ASSISTANT

# AI-Powered Internal Helpdesk Assistant Overview

An innovative solution for addressing employee queries using AI with seamless document integration and natural language processing capabilities.

DOCUMENT PROCESSING

# Document Extraction and Preparation for AI Systems

This presentation covers the steps involved in document extraction, including PDF upload, text extraction, and chunking for AI-powered internal helpdesk solutions.


VECTOR INDEXING

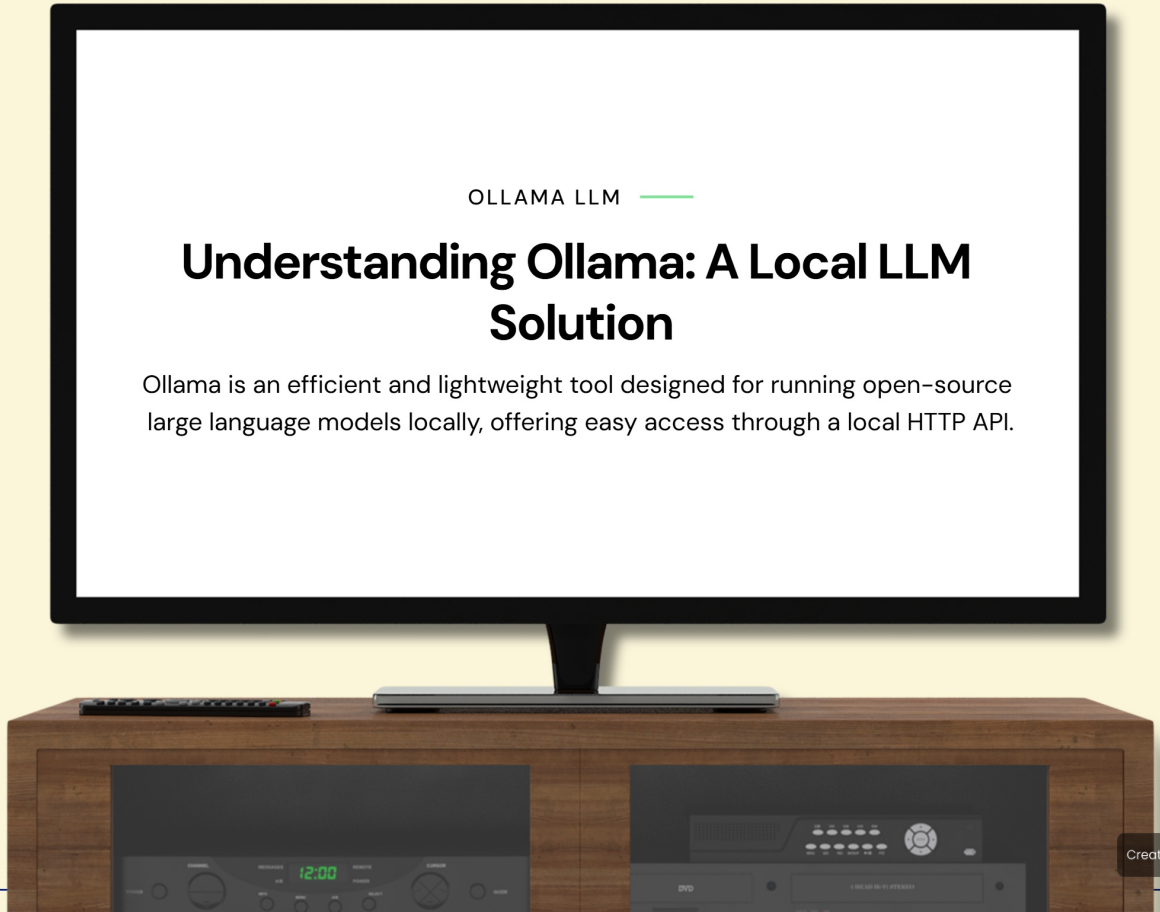
# Embedding and Vector Indexing in AI

This slide covers the process of converting text chunks into 384-dimensional embeddings, storing them in FAISS, and saving original text for retrieval.

OLLAMA LLM

# Understanding Ollama: A Local LLM Solution

Ollama is an efficient and lightweight tool designed for running open-source large language models locally, offering easy access through a local HTTP API.



# Key Advantages of Using Ollama for AI Solutions

Discover the compelling benefits of integrating Ollama for your internal helpdesk assistant.


### Local Inference: Run on user's machine

With **local inference**, Ollama processes data directly on the user's device, ensuring that no internet connection or external API keys are necessary. This enhances performance and privacy.

### Free and Open Source: No cost involved

Ollama is completely **free and open source**, allowing users to implement it without any financial constraints. There are no hidden costs or usage limits to worry about.

### Secure by Design: Protect company data

Ollama is designed with security in mind, meaning that **company data** remains on the user's system. This reduces risks associated with data breaches and enhances confidentiality.

### Easy to Use: Minimal configuration needed

Ollama is engineered for simplicity, requiring **minimal configuration** for setup. Users can quickly integrate it into their systems without complex procedures or complications.


### API Integration: Simple POST request interface

The platform offers straightforward **API integration** through a simple POST request interface, making it easy to connect with existing systems and workflows.

# Integrating Ollama into Helpdesk Systems for Enhanced Query Resolution

Leveraging AI to streamline employee support and information retrieval

### Streamlined Employee Support

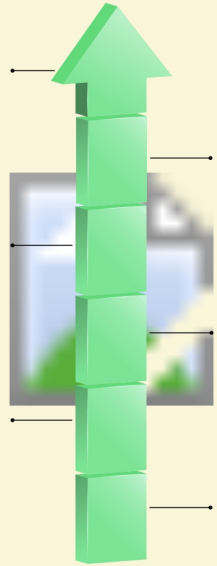
By integrating Ollama, the helpdesk system significantly improves the response time and accuracy of employee support, creating a more efficient workflow.

### User Query Processing

When a user submits a query, the system retrieves pertinent policy sections from the database. This enhances the accuracy of the responses generated by the AI assistant.

### API Call Structure

The HTTP POST request to the Ollama API includes parameters such as model and prompt. This structure allows for flexibility in generating responses based on user queries, providing tailored information.




### Answer Generation

The system utilizes the Ollama model to generate answers based on the retrieved policy sections. This process ensures that employees receive accurate and helpful information quickly.

### Response Handling

The response from the API is processed to extract the relevant information. The JSON response includes the generated text, which is then utilized in the helpdesk system for employee queries.

### Code Integration using Python

The integration with the helpdesk system is executed through a Python script that makes HTTP requests to the Ollama API. This enables seamless interaction with the AI model for generating responses.

# Understanding the Internal Processes of Ollama in Helpdesk Systems

A deep dive into the steps involved in Ollama's operation for employee queries

1

### Model Loading into RAM

The requested **model** is loaded into **RAM** to ensure quick access and processing capabilities for queries.

2

### Tokenization of Input Text

Input text is **tokenized** into manageable pieces, allowing the model to analyze and process the data effectively.

3

### Passing through Transformer Layers

The model processes tokenized data through **transformer layers** to understand **word dependencies** and context within the input.

4

### Final Response Delivery as JSON

Once generated, the answer is returned as a **JSON object**, providing a structured response for easy integration.

# Streamlined AI Workflow: From Query to Response

Efficiently addressing employee questions with AI-powered processing and integration

1

### Input: User submits a question for assistance.

The workflow initiates when a user inputs a query such as, 'What is the maximum number of sick leaves allowed?' for guidance on company policies.

2

### Backend Processing: Embedding and retrieval.

The system processes the question using **SentenceTransformer** for embedding, followed by document chunk retrieval with **FAISS** to find relevant information.

3

### Prompt Construction: Sending to language model.

A prompt is constructed based on the retrieved information and sent to **Ollama**, which is equipped to generate human-like responses to queries.

4

### Answer Generation: Providing a response.

The AI generates a detailed answer, such as 'Employees can take up to 10 sick leaves per year as per the policy,' effectively addressing the user's question.

# Let’s implement the Internal Helpdesk Assistant to revolutionize HR engagement.

Consider deploying the AI-powered Internal Helpdesk Assistant to streamline HR operations, enhance employee engagement, and ensure efficient resolution of queries. This innovative solution can significantly transform the way HR departments interact with employees, leading to improved satisfaction and productivity.