



# HR Assistant AI-Powered Policy & SOP Discovery

Automating HR Queries with  
Modern AI Models

By:

Varun Patle   Prathamesh Nimbalkar   Aditi Verma   Kanchan Bai



# Project Overview



## **AI-Powered Assistant**

Designed for HR teams and employees.



## **Quick Answers**

Addresses questions on company policies, SOPs, and HR documents.



## **Advanced AI**

Utilises semantic search and Q&A models.

# Key Features



## Semantic Search

Finds relevant documents using vector embeddings.



## AI Q&A

Natural language answers from company documents.



## Multiple Offline/Online AI Models Support

Supports Cohere, Mistral, DeepSeek, and local LLMs.



## Chat History

Stores conversations for future reference.



## Conversational Memory

Get response with context of past conversations.



## Multi Documents Support

Model supports multi Documents like PDF, Word, Text



## Modern UI

Easy-to-use web interface.



## Easy New Model Integration

New models can be easily integrated in our Code base.

# System Architecture

- **Frontend:** React (Vite) web app.
- **Backend:** Node.js (Express) with LangChain.
- **Embedding:** Hugging Face Transformers Embeddings.
- **Vector Database:** Supabase for embeddings.
- **Chats Storage:** MongoDB Atlas.
- **AI Models:**  
Online Models - Cohere, Mistral.  
Offline Models - DeepSeek, Llama.



# Supported AI Models

Cohere (cohere command-r-plus)	\$20.00 Total (5M in + 5M out)	Reliable, Good for general tasks, fast
Mistral (mistral-small)	\$62.50 Total (5M in + 5M out)	Accurate, Scalable, Enterprise-grade performance
DeepSeek (deepseek-r1:latest)	Free/Open-source, HardWare Requirements (8-16 GB RAM , 10-50 GB disk)	Privacy, full control, Zero operational costs
Llama (llama3.2)	Free/Open-source, HardWare Requirements (4-32 GB RAM , 2-180 GB disk)	No API cost, full privacy, Commercial use allowed





# Why These Models?

## **Accuracy**

State-of-the-art for language tasks.

## **Flexibility**

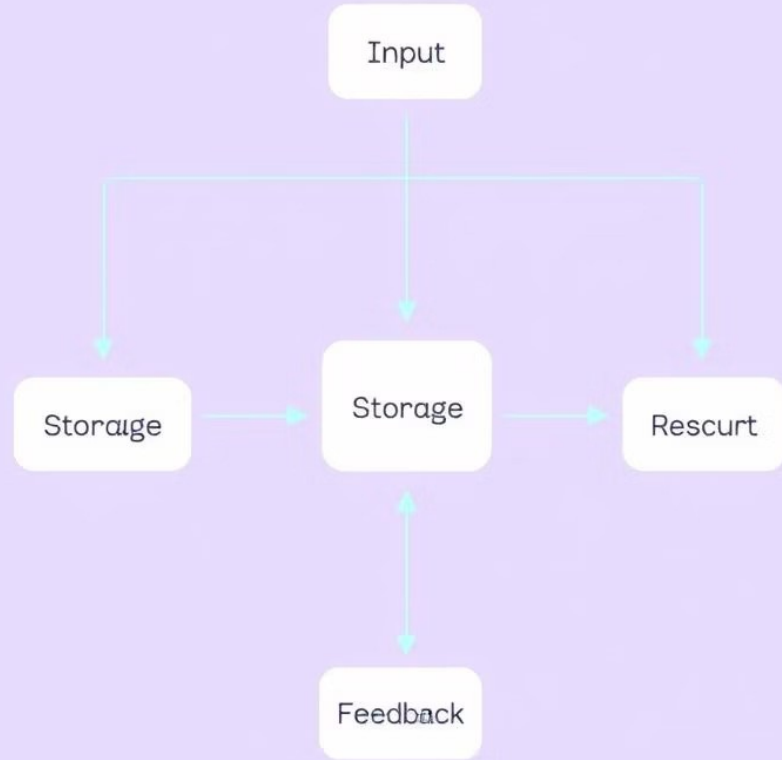
Cloud and local options available.

## **Cost Control**

Choose some open source and less costly with enterprise support.

## **Enterprise Ready**

APIs, SLAs, and open-source options.



# How It Works



## User Query

User asks a question in the web app.



## Document Search

Backend searches relevant documents using embeddings.



## AI Generates

Selected AI model generates an answer.



## Response Display

Response and sources shown to the user.

# Adding New Documents

## Place Document (PDF, Word, Text)

New Documents go into the backend9s data/ folder.

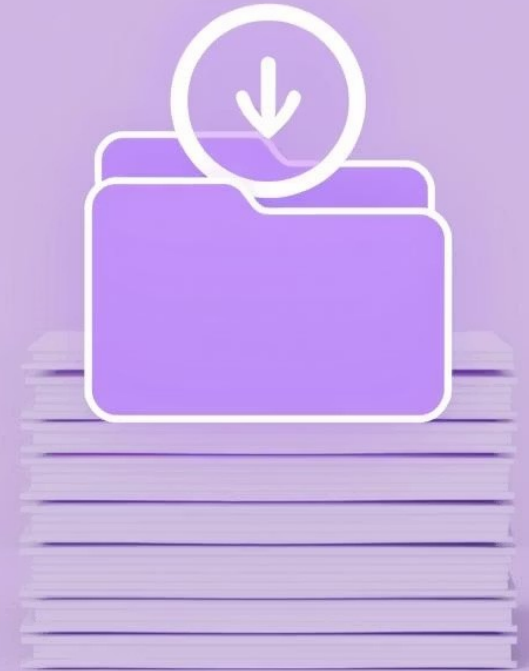
## Run Script

Execute the embedding script to update the vector database.

Refer Read Me file for script.

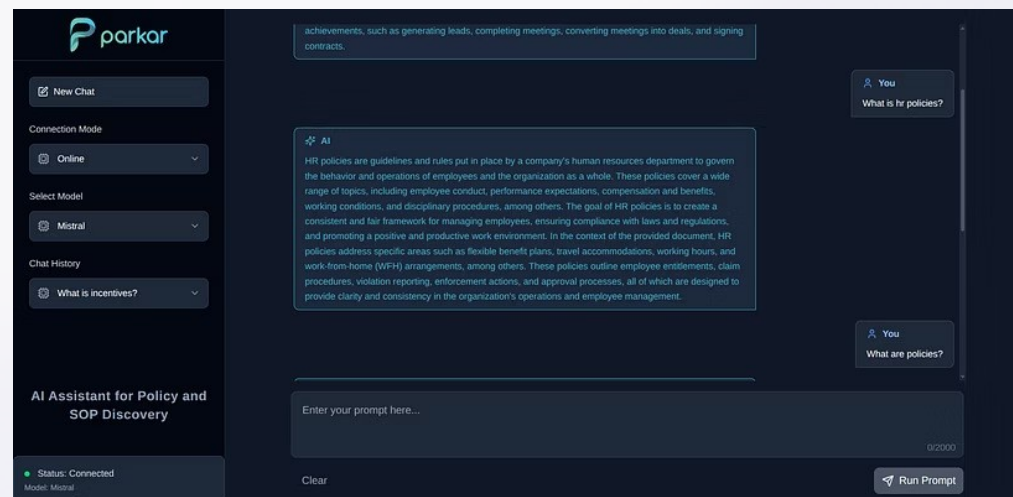
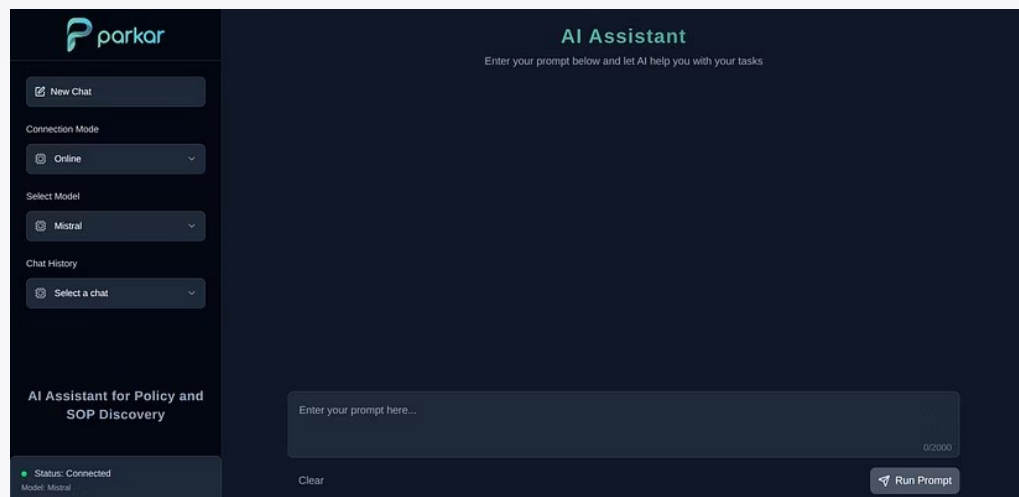
## No Code Changes

The system requires no code modifications.





# Application Screenshots:





# Future Enhancements



## More Languages

Expand language support for global use.



## HRMS/ERP Integration

Seamlessly connect with existing HR systems.



## Advanced Analytics

Implement detailed reporting and insights.



## Voice Assistant

Integrate voice interaction capabilities.