

# **EDUCATION CHATBOT FOR MANAGING BIG DATA COURSE**

Presented by Elie Niringiyimana



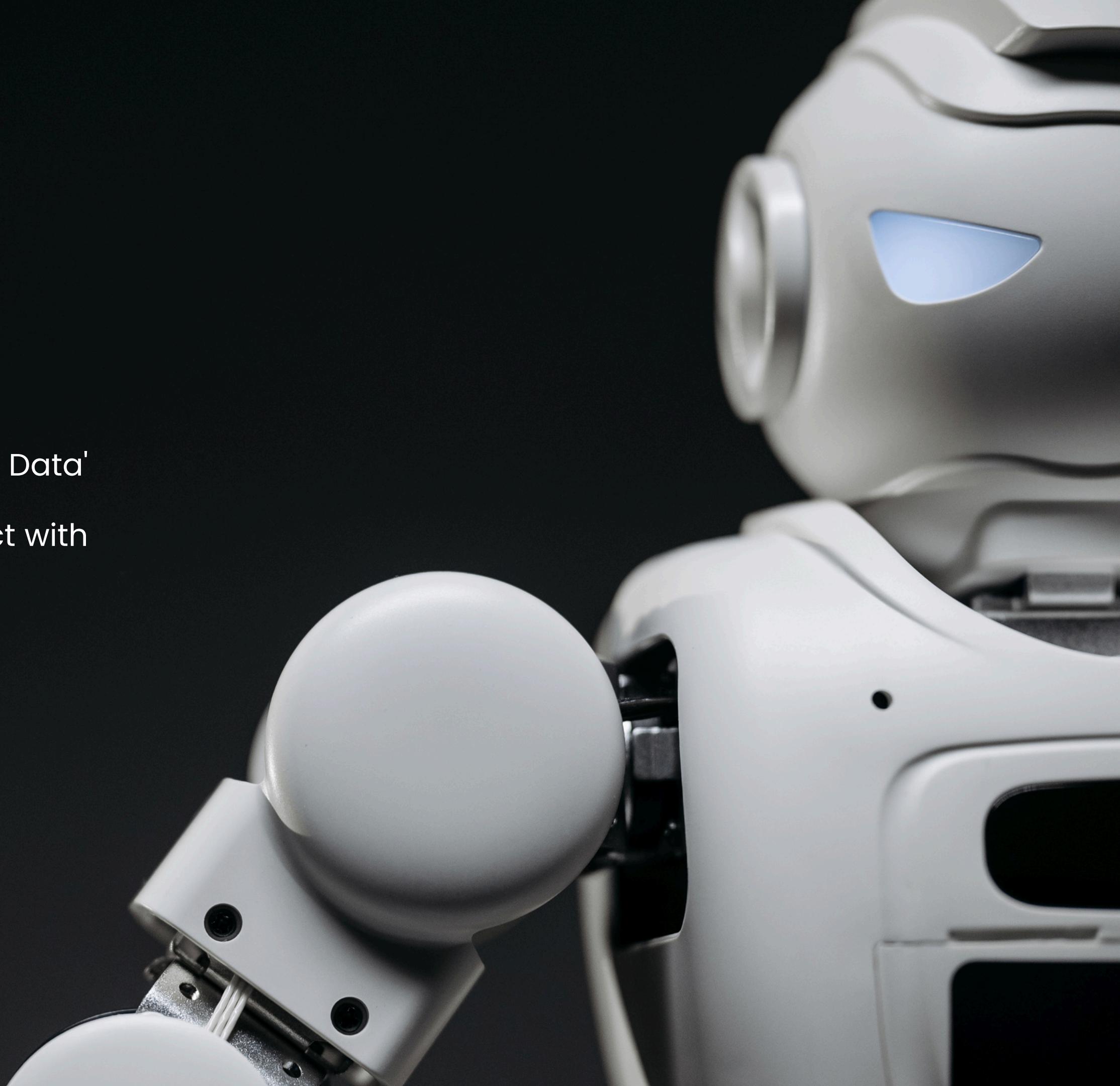
# OVERVIEW

Project Goal:

- AI-powered chatbot for the 'Managing Big Data' course.
- Professors create quizzes, students interact with the knowledge base.

Technology Stack:

OpenAI, Pinecone, Streamlit, and Ollama



# KEY FEATURES

- Extracts and Embeds Text from PDFs:  
Uses class notes to build the knowledge base
- Quiz Generation:  
Creates multiple question types (MCQ, open-ended, coding tasks)
- Context-Based Responses:  
Provides clear and accurate answers to student queries



# INSTALLATION STEPS

1

## CLONE THE REPOSITORY

- git clone:  
[https://github.com/nirin-giyimanaelie/Education\\_support\\_system.git](https://github.com/nirin-giyimanaelie/Education_support_system.git)
- cd Education\_support\_system

2

## INSTALL DEPENDENCIES

pip install -r requirements.txt

3

## SET ENVIRONMENT VARIABLES

Add API keys for OpenAI and Pinecone in the .env file

# USAGE

**Run the Streamlit App:**  
`streamlit run app.py`

**Functionality:**  
Ask questions, generate quizzes, and  
interact with course content



# EMBEDDING IMPLEMENTATION

## 1. Document Vectorization:

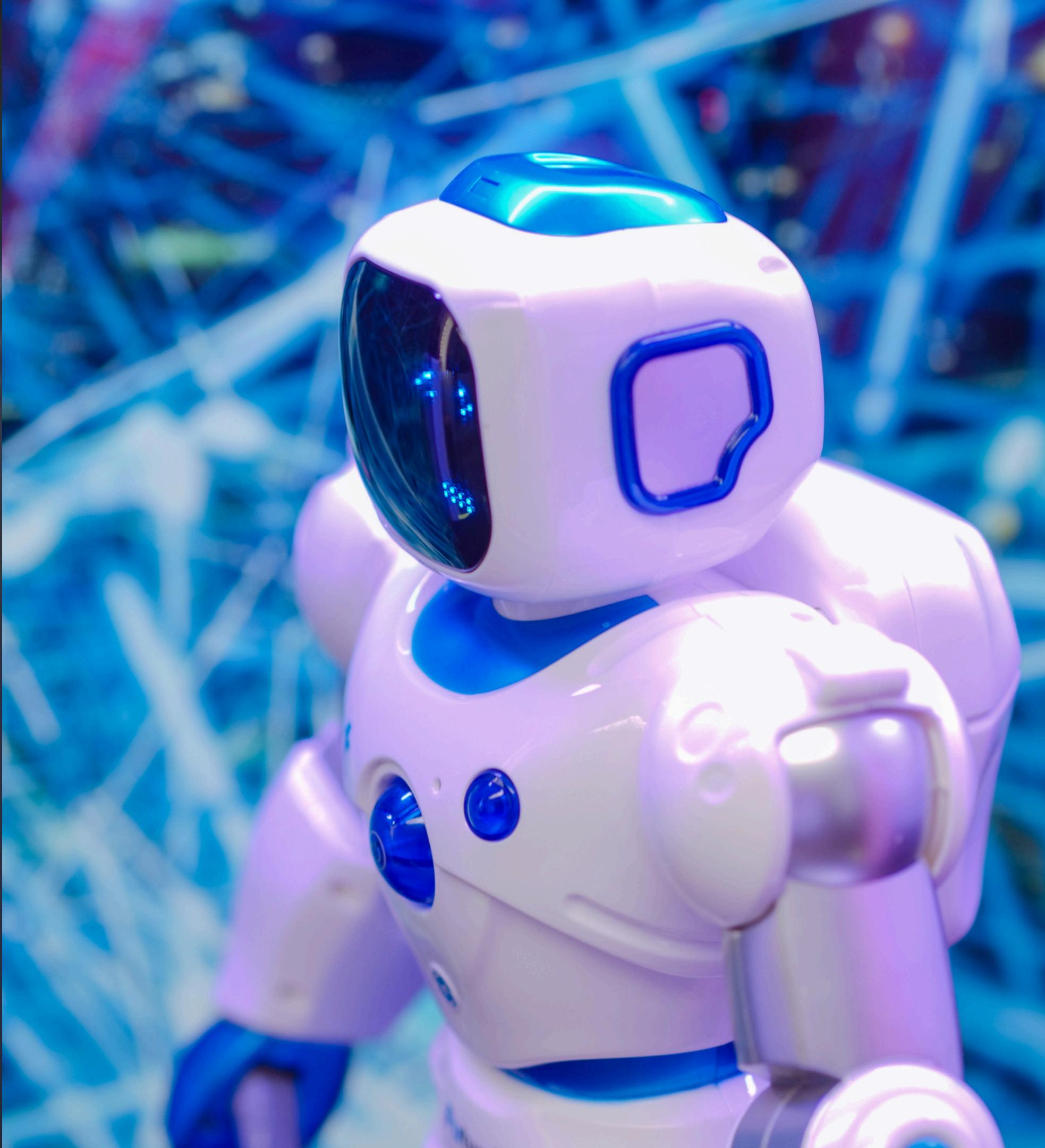
- Splitting class notes into 600-character chunks with a 100-character overlap
- Vector storage in Pinecone, including ID, vector, and metadata

## 2. Embedding Model:

- 'text-embedding-ada-002' for generating embeddings.

## 3. Augmented Queries:

- Using cosine similarity with the 5 nearest neighbors



# RESPONSE GENERATION

- Language Models Used: OpenAI's 'chatgpt-4o-latest' and Ollama's 'llama3:latest'
- Unique Aspect: Knowledge base is derived directly from class notes, ensuring relevance and accuracy



# NEXT STEPS

## **GENERATE CONTENT OF IMAGES, AND CHARTS:**

- Developing a computer vision model for image and chart identification
- Use of 'gpt-4-vision-preview' to generate content from images/charts
- Improved embedding of visual content to enhance performance

# CONCLUSION

- Innovative Solution: Combines AI and big data management for learning
- Future Enhancements: Generate contents of visual content for comprehensive quizzes and answers
- Impact: A unique study tool aligned with course-specific materials.



# DEMO

**Hosted App:** <https://managingbigdata.streamlit.app/>

**Demo:**

<https://drive.google.com/file/d/1SzNaON0vF4oo8VLDR7ret5gxR2DagSM/view?usp=sharing>

# THANK YOU

