

**HW#07**  
**SPECIFICATIONS FOR BOT**

**Team:**  
**Anjaneya Padwal**  
**Jill Karia**  
**Sudhanshu Pawar**  
**Aakanksha Maheshwari**

| <b>Topic</b>                     | <b>Page Number</b> |
|----------------------------------|--------------------|
| Challenges                       | 1                  |
| Concept and Purpose              | 1                  |
| Bot Creation                     | 2                  |
| Data Sources                     | 7                  |
| Architecture                     | 8                  |
| Bot's Usefulness                 | 10                 |
| Comparison with Bot Architecture | 11                 |

## **Bot Name: Hepha**

### **Challenges:**

#### **Coursework**

- Being familiar with a new Programming Language (R Language) for the course
- An LLM can only help in providing code and after arduous prompts can help someone understand the thought process behind the code, the reasoning for using certain R Functions and translating rough ideas and workflows into R for a dataset was a challenge

#### **IST687 Project**

- The project groups largely consisted of 4-5 students and managing the whole project as well as allocating tasks among the students was a tough task in the Project
- Another point here was understanding how the project should be approached, from data integration to data cleaning, which models should be determined, and how the results of each model impacted our solutions.

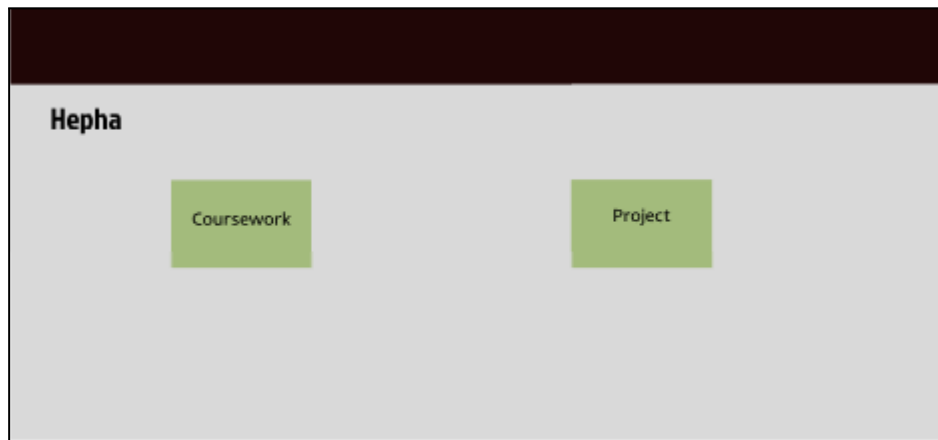
### **Concept and Purpose:**

- The purpose of the Bot is to augment the efforts of the Instructor and Course Assistants to provide real-time, on-demand, and correct help to students to excel in IST687 and take key outcomes essential in their careers moving forward
- This bot will be available on Blackboard and will have two modes viz: Coursework and Project
- The Coursework mode will allow students to interact with the bot and resolve their queries regarding coursework, help with homework, and thought processes behind several steps, and get better at using the various libraries of R needed for the course
- The Project mode will aid the students in distributing and managing tasks augmenting the efforts of a Student Project Manager essentially, and assisting in resolving queries as well as validating the Project Approach when it comes to the Final Project

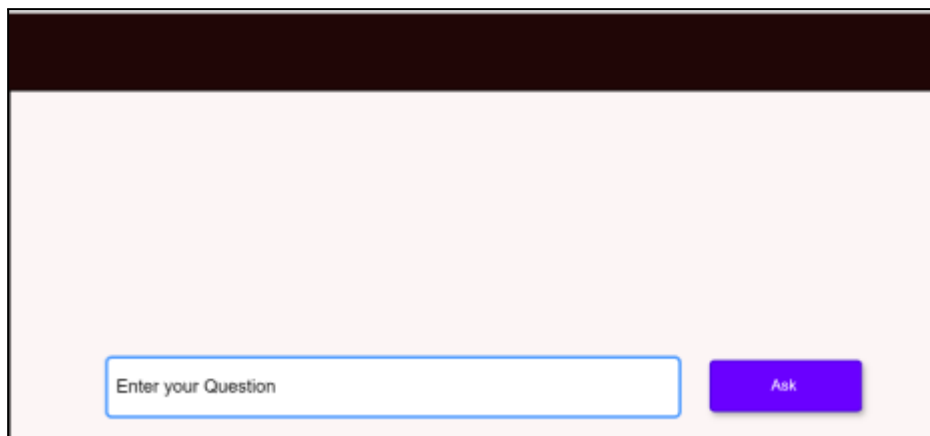
## Bot Creation:

### Coursework

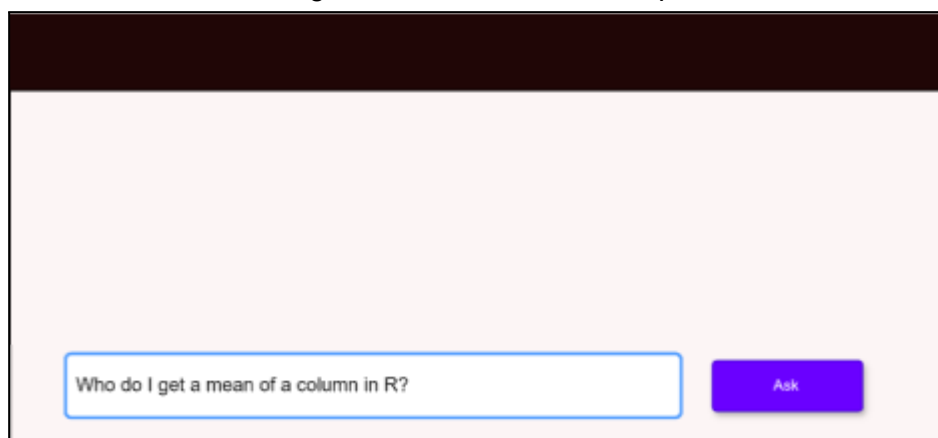
- The student opens the Bot via the Blackboard Account and clicks on the “Coursework” button as shown below



- After clicking on the “Coursework” button, a chat interface as shown below will be displayed to the user where the user can ask a question related to the coursework



- This question will first be classified as a type of question and then connecting to the relevant Knowledge Base an answer will be provided to the user



- Sample Prompts/ Questions and their types are listed below:

| Sample Question   | Question Type                 |
|---|-------------------------------|
| Is HW 6 compulsory or Optional?                                     | Coursework -Related           |
| How do I get the Mean of a column of a data frame?                  | Familiarity with R            |
| How do I get around with the Data Cleaning process for HW3?         | Thought Process around IDS    |
| How do I resolve the error "Error in ...: could not find function"? | Troubleshooting Problems in R |
| My RMSE score is 0.68, what can I do to improve the model?          | Thought Process around IDS    |

- Based on the Question, the bot will fetch the answers from the Knowledge Base and then provide an answer to the users

How do I get a mean of a column in R ?

Ask

How do I get a mean of a column in R ?

You can use the command in format `mean(dataframe$column)`. For eg. `mean(df$Sales)`

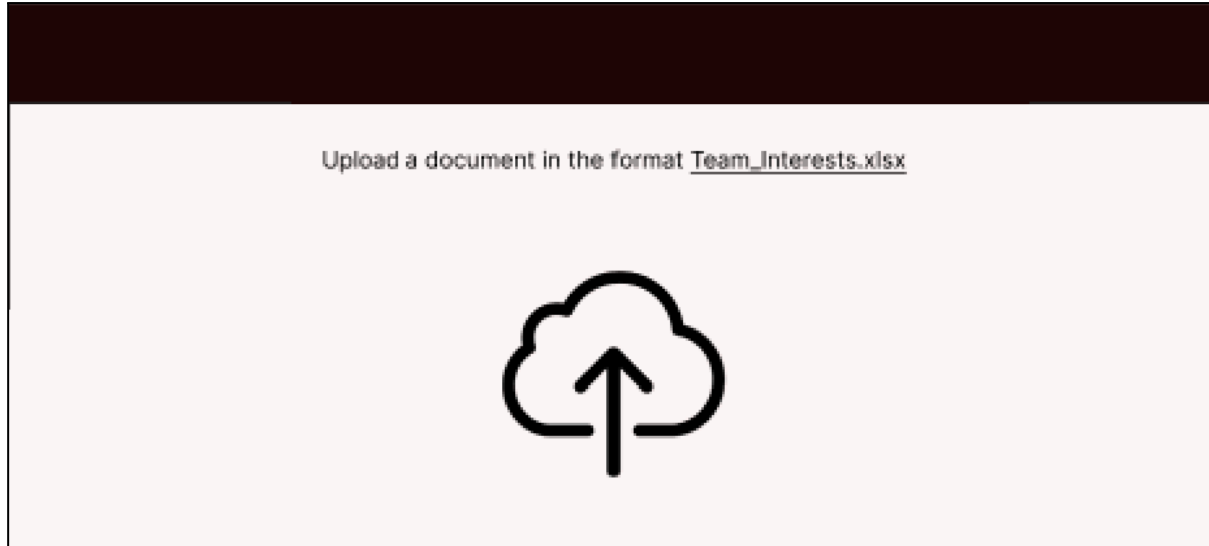
Ask

- The Knowledge will consist of databases as shown here
- The chatbot will have a memory that can support up to 20 chat messages post which the chat will be reset
- A question can consist of 700 characters post and a warning will be displayed to the user to summarize the text

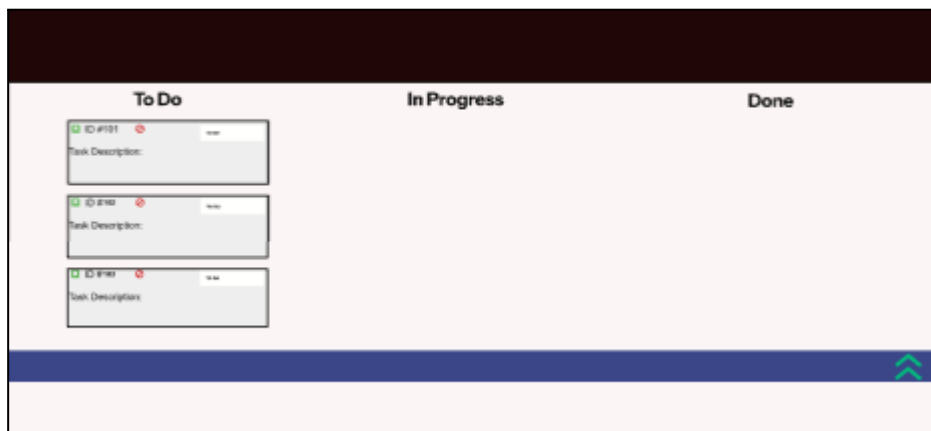
- An edge-case handling mechanism will be in place to handle any out-of-scope questions like “Which is the easiest way to get to Wescott?”

## Project

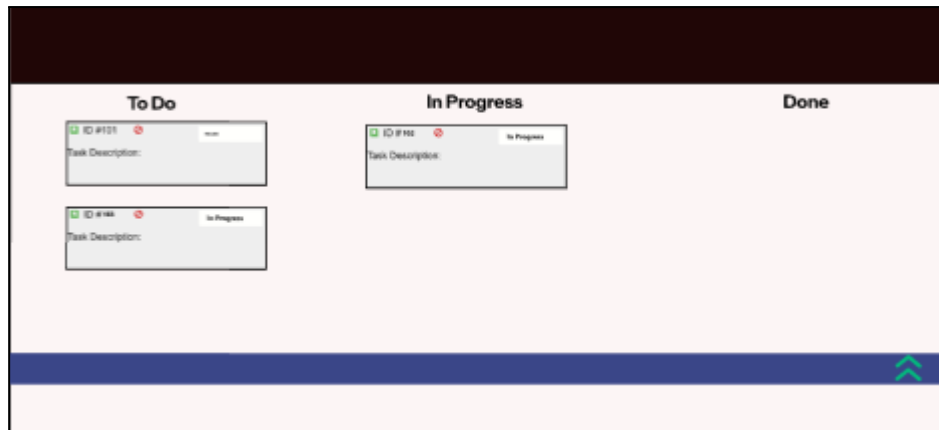
- When the user clicks on the Project tab for the first time, the user is asked to upload an Excel in the format of “Team Member: Areas of Interest”



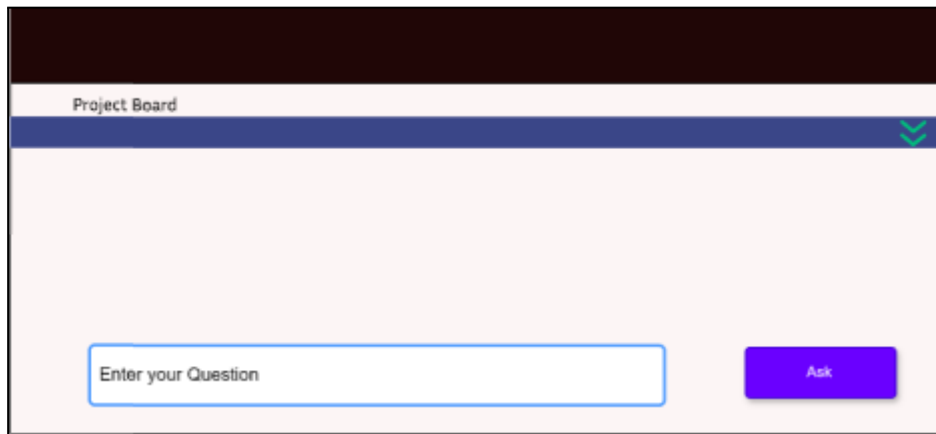
- Here, the Areas of Interest can include tasks such as Data Cleaning, Visualization, Machine Learning Modelling, and Storytelling.
- Based on these topics, the Bot will create a Scrum Board where the larger Project will be distributed into several tasks



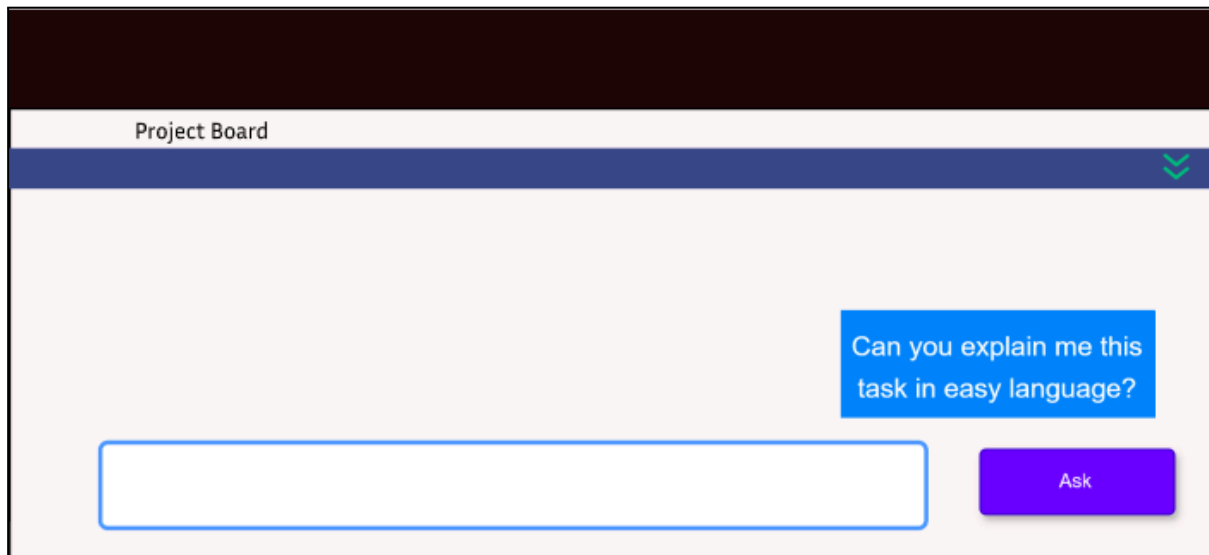
- These tasks will be in the KDD format wherein each step will be broken into tasks and written in the form of tickets
- Each ticket will have a Title, the Person Allowed to, and the description of the task in the ticket
- The tickets will be divided into three buckets viz: To do, In the process, and Done and will be arranged in a particular chronology horizontally which will be reflected by the Ticket Number
- Once the user picks their respective ticket and transfers it from To do to In process, the bot is informed about the change in status of the ticket

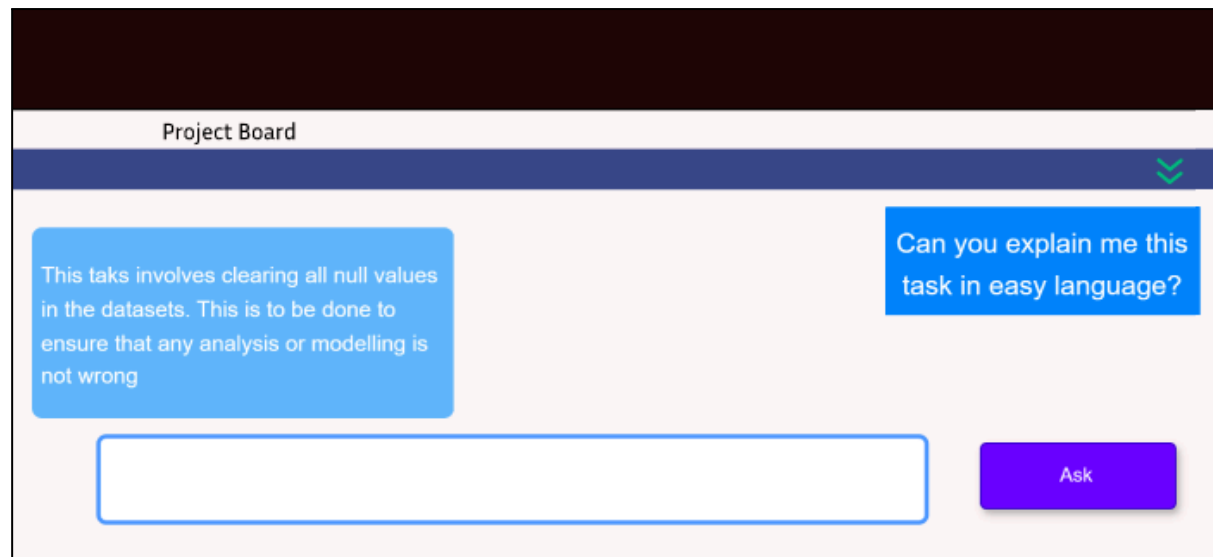


- This Project Management Board can be toggled with an accordion button clicking which minimizes the board
- Once the board is minimized, the user will see a chatbot with an interface similar to the first one



- The bot will be informed of the Project Stage and will answer questions specific to that particular project stage only





- Sample Questions that the Chatbot can answer per process

| Sample Questions  | Stage of Data Science Project |
|---|-------------------------------|
| How do we aggregate the data to cover all houses? Should we group using County ID or first predict for one house?           | Project Thought Process       |
| How do I eliminate all Null Values in the Latitude Longitude Column?  | Data Cleaning                 |
| Can I join the weather and housing datasets using joins and the same columns?   | Data Integration              |
| Facing "Error: Unhandled Exception: Child Task 71390052 error: Unhandled Exception: Execution halted". Can you help resolve | Deployment                    |
| What visualizations should be chosen for the weather data set?  | Data Visualization            |
| Should I aggregate the housing dataset to incorporate the volume of the house?  | Feature Engineering           |
| Can Arima be used if the Linear Regression RMSE Score is 0.67   | Machine Learning Modelling    |

- The bot trained over the questions in these stages will be able to answer these questions effectively essentially augmenting the work of a Project Manager
- The conversation limit for this bot will be 35 messages given the use case of this Bot post which the chat will be reset

## **Data Sources:**

### **Coursework**

- The datasets needed for the coursework would be viz:
  - IST 687 Coursework and Course Manual Data essentially comprising details on the course, assignments, and logistics of the course
  - Masked Homework and Lab Solutions
  - R Manual for IST687
  - Troubleshooting Guide

### **Projects**

- The datasets needed for the Project Bit are viz:
  - Team Members and Areas of Interest Data
  - Templates for Kanban, Agile, and KDD Project Management
  - Stage-based Information and Process Flows of the Project
  - Error handling and Process Flow
  - Runtime Troubleshoot Guide



## Architecture of the Hepha Bot

The architecture of the Hepha bot is designed to facilitate seamless interactions between students and the system within the Blackboard environment, supporting two distinct operational modes: Coursework and Project. This architecture is built upon several core components that ensure responsiveness, scalability, and effective assistance.

### User Interface (UI)

- **Modular Chat Interface:** The bot features a user-friendly chat interface that adapts to the mode selected (Coursework or Project). This modular design allows users to interact easily, ensuring accessibility and engagement.
- **Dynamic Display:** Depending on user interactions, the UI dynamically displays relevant information and options, such as task lists or help topics, enhancing user experience.

### Bot Logic and Knowledge Base

- **Natural Language Processing (NLP):** The bot employs NLP algorithms to interpret user queries, enabling it to understand and classify questions accurately based on context.

#### Knowledge Base Architecture:

- **Structured Databases:** Information is organized in a structured manner, with categories for R functions, troubleshooting techniques, data integration strategies, and project management best practices.
- **Real-time Retrieval System:** The bot connects to a vector database that allows for quick retrieval of relevant information based on similarity searches, ensuring timely responses to student queries.

### Memory and Interaction Management

- **Session Memory:** The bot maintains a session memory that supports up to 20 interactions, allowing for context retention and a coherent conversation flow. After reaching the limit, the session is reset to maintain performance.
- **Error and Edge Case Handling:** Robust mechanisms are in place to manage out-of-scope questions effectively. For instance, queries that do not align with the bot's purpose prompt the user with helpful responses or redirections.

### Project Management Board

#### Scrum Board Implementation:

- **Task Allocation and Tracking:** Upon uploading team member interests, the bot generates a Scrum board that breaks down the project into manageable tasks (tickets) in a KDD (Knowledge Discovery in Databases) format.

- **Task Management Features:** Each ticket includes a title, assigned team member, and detailed task description, organized into three states: To Do, In Progress, and Done.
- **Real-Time Status Updates:** As users move tasks between states, the bot updates the Scrum board in real-time, ensuring all team members stay informed about project progress.

### User Interaction Flow

- **Toggle Functionality:** Users can seamlessly switch between the Scrum board and the chat interface using an accordion feature, allowing them to manage their projects while addressing coursework queries effectively.
- **Contextual Awareness:** The bot retains information about the current project stage and adapts its responses accordingly, ensuring relevance and precision in the assistance provided.

### Use of Vector Databases

- **Information Representation:** Concepts within the Knowledge Base, such as R functions and troubleshooting tips, are represented as high-dimensional vectors. This allows for semantic understanding and similarity matching between user queries and stored knowledge.
- **Efficient Query Processing:** When a user asks a question, the bot converts the query into a vector and retrieves the most relevant responses from the database using similarity metrics, significantly reducing response time and improving accuracy.
- **Scalability:** The use of vector databases allows for easy scaling of the Knowledge Base, accommodating new information and updates without significant architectural changes.

### Use of Functions

- **Query Processing Functions:** These functions handle the initial processing of incoming questions, including NLP tasks for understanding intent and classifying the question type.
- **Response Retrieval Functions:** Responsible for fetching answers from the Knowledge Base, these functions query the vector database to find relevant responses based on classified question types.
- **Project Management Functions:** These include functionalities for creating, updating, and managing tasks on the Scrum board, ensuring smooth transitions between task states and accurate status reporting.
- **User Interaction Functions:** Functions that manage the chat interface, allowing for dynamic updates and feedback based on user input, enhancing user experience and engagement.

## **Hepha's Usefulness**

### **Supporting the Learning Objectives of the Course:**

- Hepha is designed to enhance the learning experience in IST 687 by acting as a supplement to the Instructor and Course Assistants.
- The bot aligns with the key learning objectives by assisting students in grasping essential R programming concepts, data analysis techniques, and project management skills.
- Its "Coursework" mode focuses on resolving course-specific queries, helping students develop a better understanding of R libraries, troubleshooting issues, and applying the correct thought process in solving homework problems.
- By providing timely feedback and contextual guidance, the bot enables students to apply theoretical knowledge more effectively to real-world problems, promoting a deeper understanding of course content.

### **Impact on Student Engagement, Understanding, and Performance:**

- The bot's presence offers real-time, on-demand assistance, which is crucial for maintaining student engagement. It allows students to address their coursework issues independently, fostering a sense of ownership over their learning process.
- By interacting with Hepha, students can explore difficult concepts and R functions, leading to an improved understanding of data cleaning, integration, and model evaluation, which are critical skills in the IST 687 course.
- Additionally, Hepha's structured project management tools help students stay organized, ensuring timely task completion, which ultimately boosts their project performance and overall course grades.

## **Comparison with Current Bot Architectures**

### **Comparing Hepha with Existing Bots like ChatGPT**

- Hepha's architecture is specifically tailored to meet the needs of students in the IST 687 course, whereas bots like ChatGPT are general-purpose AI models designed for broader conversations.
- While ChatGPT can assist with generic questions, it often requires detailed prompts and additional context to provide accurate answers related to specific coursework or project-related queries.
- Hepha, on the other hand, is designed with a specialized knowledge base built around IST 687, allowing it to quickly fetch targeted answers from the course materials, R programming resources, and project management guidelines.

### **Advantages and Suitability of Hepha's Design for IST 687**

- Hepha's architecture offers several advantages for IST 687 students. Its ability to provide context-specific answers based on course-related data makes it far more suitable than general-purpose bots.
- For instance, Hepha can validate project approaches by helping students structure their tasks according to the KDD process and assist them in resolving R programming errors or conceptual misunderstandings.
- The integration of a project management board further augments Hepha's usefulness by helping students manage their group projects effectively, something that generic bots like ChatGPT are not equipped to handle.
- This specialized design ensures that students not only receive accurate technical help but also stay organized and on track with their projects.