

# A21.ai Assignment

AI based text summarization web application

Sinjini Ghosh

## **AI based text summarization web application**

This is an AI-based web application that summarizes the text entered by the user either manually or through a PDF file. It is built using the python web framework: Flask.

### **Functionality**

The application provides two options for users to input text data for summarization:

#### **A. Manual Text Entry:**

Users can manually enter text in the provided text area. The AI model will then generate a concise summary based on the input text. For optimal results, it is recommended that the entered text contains between 30 to 1000 words.

#### **B. PDF File Upload:**

Alternatively, users can upload a PDF file containing textual data. The model will process the content and generate a summary. Similar to manual entry, it is advisable to have a PDF file with a word count ranging from 30 to 1000 for optimal summarization.

Additionally, the application has the following functionalities:

#### **User Registration:**

Users can register on the website by providing a unique username and password.

#### **User Login:**

After registration, users can log in using their registered username and password.

#### **Summarization History:**

The summarization history for each registered user is stored and conveniently displayed on the left side of the page, allowing users to review their interaction history.

## Database to store user data

The application utilizes an SQLite database with two tables:

A. User Table:

Stores user data, including unique user IDs, usernames, and passwords.

B. Summarization History Table:

Records the interaction history of users, preserving their summarization activities.

## Link to the github repository:

<https://github.com/Sinjini2508/A21.ai-Assignment>

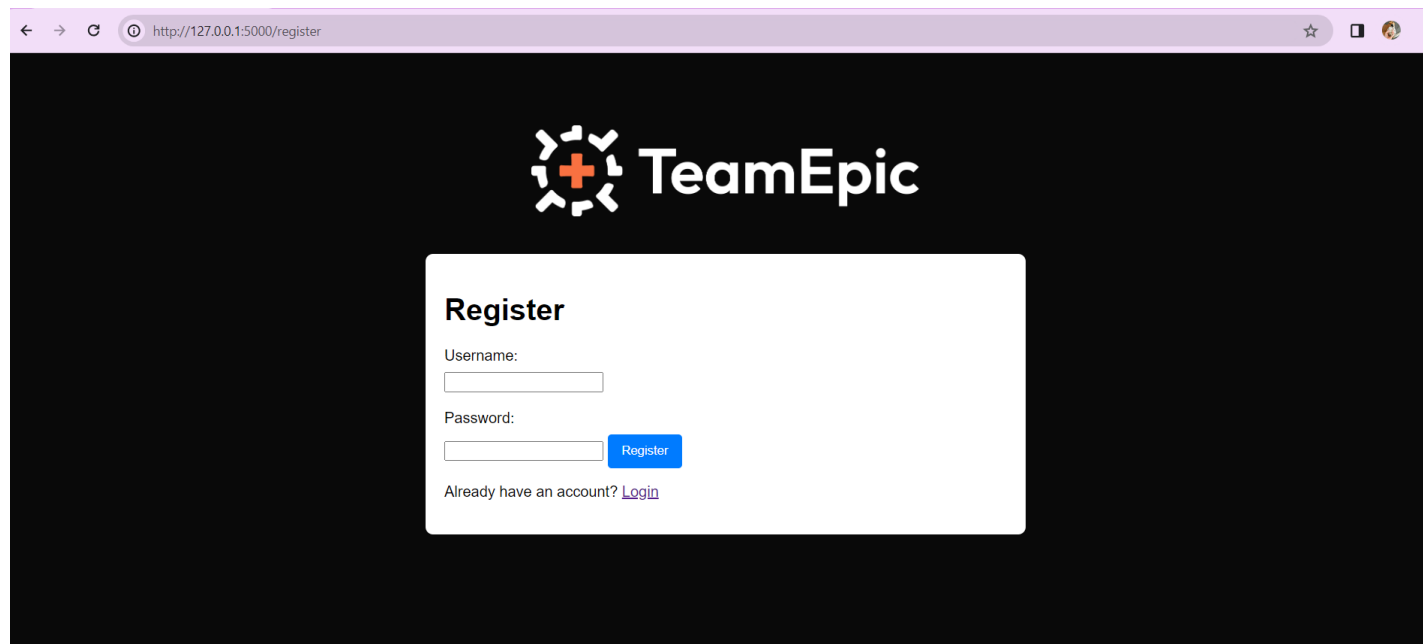
## Steps to run the application locally.

1. Clone the repository: git clone <https://github.com/Sinjini2508/A21.ai-Assignment>
2. Navigate to the project directory
3. Run the app.py file
4. Access the application in your web browser at <http://127.0.0.1:5000>
5. Register a new account on the registration page or login using the existing account:  
[Username: user1](#)  
[Password: passwordfortesting](#)

**Demo Video:** A demo video without a voiceover is included in the zipped folder.

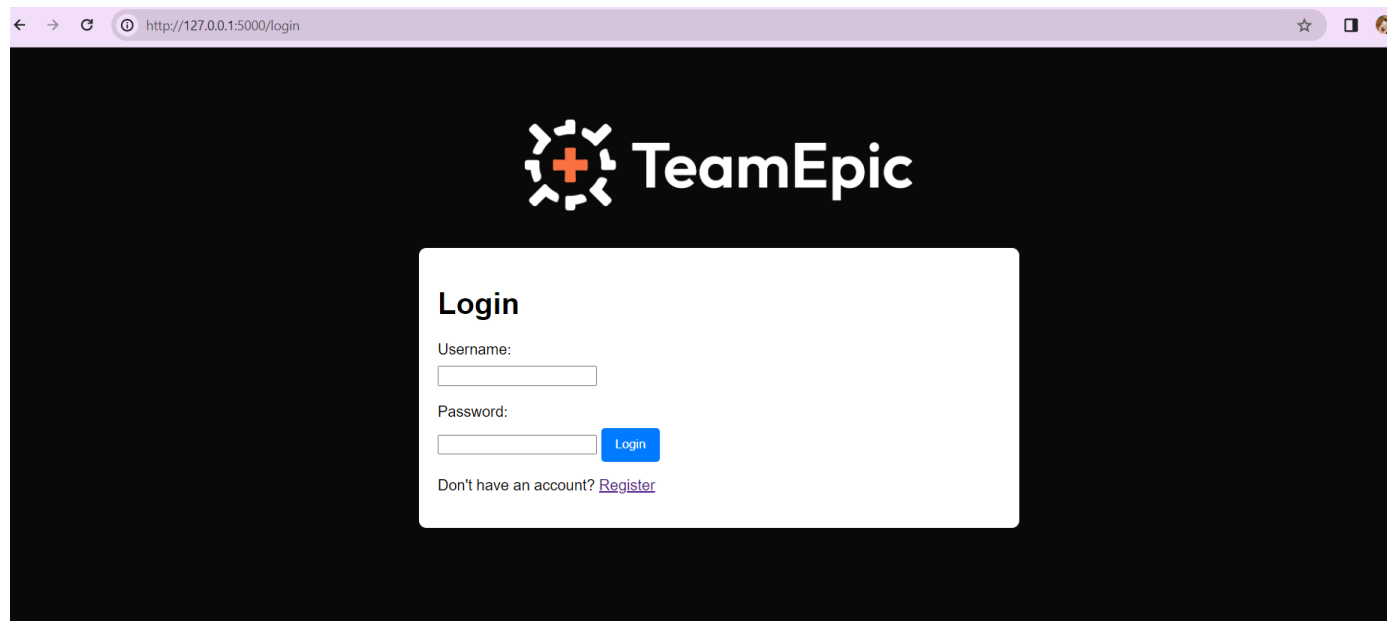
## Screenshots from the application

### Registration page



A screenshot of a web browser showing the registration page for TeamEpic. The browser's address bar displays 'http://127.0.0.1:5000/register'. The page has a dark background with the TeamEpic logo at the top center. Below the logo is a white rectangular form titled 'Register'. Inside the form, there are two input fields: 'Username:' and 'Password:'. To the right of the password field is a blue 'Register' button. Below the password field, there is a link that says 'Already have an account? [Login](#)'.

### Login Page



A screenshot of a web browser showing the login page for TeamEpic. The browser's address bar displays 'http://127.0.0.1:5000/login'. The page has a dark background with the TeamEpic logo at the top center. Below the logo is a white rectangular form titled 'Login'. Inside the form, there are two input fields: 'Username:' and 'Password:'. To the right of the password field is a blue 'Login' button. Below the password field, there is a link that says 'Don't have an account? [Register](#)'.

## User Input Page (Left panel for interaction history, right panel for entering text)

The screenshot shows a web browser at <http://127.0.0.1:5000/home>. The page is divided into two main sections. The left section, titled "Summarization History", contains two entries. The first entry is for "Indigo" by South Korean rapper RM of BTS, released on December 2, 2022, through Big Hit Music. It serves as a documentation or archive of his late twenties, comprising 10 tracks, including appearances by Erykah Badu, Anderson .Paak, Tablo of Epik High, Kim Sa-wol, Paul Blanco, Mahalia, Colde, Youjeen of Cherry Filter, and Park Ji-yoon. The ninth track, "Wild Flower", a collaboration with Youjeen, was released alongside the album as its lead single, together with an accompanying music video. The album peaked at number two in South Korea; number three in Lithuania, Portugal, and the United States—the highest-charting album by a Korean soloist on the Billboard 200 at the time;[A] and number four in Japan. Indigo was certified double platinum by the Korea Music Content Association in January 2023 and has sold over 800,000 copies domestically as of July 2023, and over 100,000 units in the US. The second entry is for "Anette Moldvaer states that a flat white consists of a double espresso (50 ml/1.5 fl oz) and about 130 ml (4 fl oz) of steamed milk with a 5 mm (0.25 inch) layer of microfoam.[1] According to a survey of industry". The right section, titled "Text Summarization", features the TeamEpic logo and two options. Option 1: "Please enter your text below (minimum 30 words and maximum 1000 words for optimal results):". It includes a text input field and a "Submit" button. Option 2: "Please upload a PDF file (minimum 30 words and maximum 1000 words for optimal results):". It includes a "Choose File" button, a "No file chosen" status, and an "Upload and summarize" button.

## The summary result page

The screenshot shows a web browser at <http://127.0.0.1:5000/summarize>. The page features the TeamEpic logo and a section titled "Summary Result". Under "Original Text:", it displays a paragraph about Hedy Lamarr: "Hedy Lamarr was an Austrian-American actress and inventor who pioneered the technology that would one day form the basis for today's WiFi, GPS, and Bluetooth communication systems. As a natural beauty seen widely on the big screen in films like Samson and Delilah and White Cargo, society has long ignored her inventive genius. Lamarr was originally Hedwig Eva Kiesler, born in Vienna, Austria on November 9th, 1914 into a well-to-do Jewish family. An only child, Lamarr received a great deal of attention from her father, a bank director and curious man, who inspired her to look at the world with open eyes. He would often take her for long walks where he would discuss the inner-workings of different machines, like the printing press or street cars. These conversations guided Lamarr's thinking and at only 5 years of age, she could be found taking apart and reassembling her music box to understand how the machine operated. Meanwhile, Lamarr's mother was a concert pianist and introduced her to the arts, placing her in both ballet and piano lessons from a young age." Under "Summary:", it displays a shorter version of the same text: "Hedy Lamarr was an Austrian-American actress and inventor who pioneered the technology that would one day form the basis for today's WiFi, GPS, and Bluetooth communication systems . As a natural beauty seen widely on the big screen in films like Samson and Delilah and White Cargo, society has long ignored her inventive genius ."

## Tech Stack

Programming language: Python

Scripting language: HTML and CSS

Python Web Framework: Flask

LLM: Falconsai/text\_summarization

LLM Framework: Langchain

Vector Database: FAISS

SQL Database: SQLite

## Workflow Diagram

