Interim Presentation

A Regulated and Secure Mental Support Chatbot Featuring a User- Large Language Model (LLM)-User Sandwich Architecture

By CHEN Xingyi, CHEUNG Kiki, LEE Tsz Shan Jessica, WU Yuxuan, YAN Wenhao
Mentor: Professor YIU Siu Ming
Date of Presentation: 1st June, 2024

Presentation Outline

1. **Follow-up on Progress Updates 2**- Literature Review & Research Gap
2. **Preliminary Results**- Current Project Status
3. **Schedule** for the Rest of the Project

1. Literature Review & Research Gap

Through the analysis of existing research to identify research gap thereby emphasize the significance and relevance of our project.

1. Current Mental Health Situation

therapists. The numbers of professionals working in the psychiatric stream in HA is in **Appendix 1**. In 2022-2023 (projection as of 31 December 2022), a total of around 293 000 psychiatric patients were treated in HA (including inpatients, patients at specialist outpatient clinics and day hospitals), with the number of new cases at psychiatric specialist outpatient clinics and median waiting time in the same year in **Appendix 2**. The expenditure for providing mental health services by HA in 2022-2023 was $6,086 million (revised estimate).

Link: https://www.legco.gov.hk/yr2023/english/panels/hs/papers/hs20231117cb4-977-5-e.pdf

1) Increased investment in the field of mental health services by the Hong Kong Hospital Authority revealed in the 2022-2023 Background brief prepared by the Legislative Council Secretariat for the meeting on 17 November 2023

1. Current Mental Health Situation

Hong Kong's overall mental health and well-being levels have been of concern for a while now, and the COVID-19 situation has only highlighted the extent of the crisis and the urgency to do more to help. Mind HK's research showed:

- 38% (37.7%) of respondents said their mental health has gotten worse since January 2022
- Almost 56% (55.6%) of respondents scored lower than 52 out of 100 on the World Health Organisation 5 (WHO – 5) Well-Being Index, indicating "poor" overall mental well-being
- Nearly half (49.4%) of the respondents showed symptoms of mild to severe depression, and 19.2% showed moderate to severe symptoms of depression, according to the Patient Health Questionnaire – 9 (PHQ – 9)
- 41.3% surveyed showed symptoms of mild to severe anxiety, and 13.7% showed symptoms of moderate to severe anxiety, according to the General Anxiety Disorder – 7 Questionnaire (GAD – 7)

These figures clearly show that public mental health in Hong Kong is of significant concern and should be addressed as a top priority for society.

mindhk 香港心聆

A mental health charity (S88 registered – 91/16471) in Hong Kong.

Link: https://www.mind.org.hk/press-releases/mind-hk-survey-reveals-hong-kong-citizens-worsening-state-of-mental-health-during-the-fifth-wave-of-the-covid-19-pandemic/

2) Nearly half of the surveyed respondents conducted by Mind HK displayed symptoms of various degree of depression and other mental health concerning issues

1. Current Mental Health Situation

In the past 12 months (01 April 2023 - 31 March 2024), there were 51,372 new case bookings in Psychiatry Specialist Out-patient Clinics, with waiting time as follows:

	New case bookings(2)		Waiting Time							
	Number	Percentage		Hong Kong East	Hong Kong West	Kowloon Central	Kowloon East	Kowloon West	New Territories East	New Territories West
Urgent Case	2,607	5%	Median	1 week	1 week	Less than 1 week	1 week	Less than 1 week	1 week	1 week
Semi-urgent Case	9,592	19%	Median	3 weeks	4 weeks	3 weeks	3 weeks	3 weeks	4 weeks	3 weeks
Stable Case	39,170	76%	Median	26 weeks	40 weeks	20 weeks	48 weeks	26 weeks	76 weeks	35 weeks
			Longest(1)	86 weeks	83 weeks	81 weeks	91 weeks	95 weeks	99 weeks	95 weeks

(For reference only)
Date of Next Update: 31 July 2024

Notes:
(1) The longest (90th percentile) waiting time implies that appointments are earlier than the indicated time in 90% of the new case bookings
(2) Excluding cases pending for triage
(3) With effect from 1 October 2022, the waiting time for new case booking at Specialist Out-patient Clinics has incorporated integrated clinics

醫院管理局
HOSPITAL AUTHORITY

Link:
https://www.ha.org.hk/visitor/sopc_waiting_time.asp?id=7&lang=ENG

3. There is a limited relevant and accessible welfare resources available

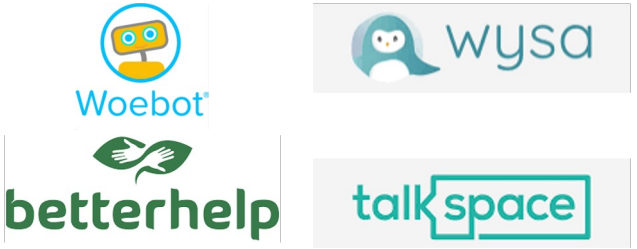
2. Existing Mental Health Supporting Technology & Traditional Mental Health Interventions

1) **Psychoanalysis**- improving negative or undesired behaviors through acquiring new techniques to create sustainable changes
2) **Behaviour therapy**- a type of behavioral therapy with the emphasis on the identification of the relationship between thoughts and feelings, leading to a healthier response by the individual
3) **Cognitive therapy**- a type of behavioral therapy with the emphasis on the identification of the relationship between thoughts and feelings, leading to a healthier response by the individual
4) **Humanistic therapy-** examination of the individual's value of the world and themselves in the world to help the individual to recognize their strength and responsibility to evolve into a fuller version of themselves
5) **Integrative therapy**- a combination of the above

- Long history of showing their effectiveness
- Heavy reliance on in-person consultations
- Therapy alone is inadequate to address the rapidly growing demand for convenient, cost-effective, and scalable mental health services

2. Existing Mental Health Supporting Technology & Traditional Mental Health Interventions

3 major AI tools used in the current mental healthcare system as a supporting role:



1. Chatbot-based therapy- the use of chatbots/ conversational agents as a form of therapy or support for mental health, mainly in the form of cognitive-behavioural therapy

2. Emotional Health Apps- designed to support individuals in managing and improving their emotional well-being, typically typically offering features, such as mood tracking, guided meditations, stress reduction exercises, journaling prompts, and self-help resources

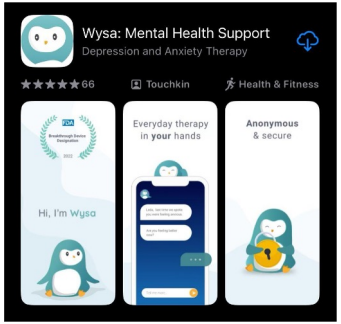
3. Smart Mental Health tools- employ advanced technologies and data analysis to support mental health- such as smartwatches or fitness trackers, that monitor physiological indicators like heart rate or sleep patterns to provide insights into an individual's mental well-being

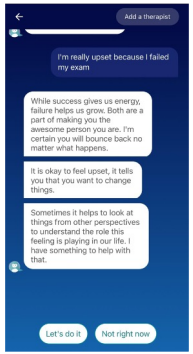
https://www.sciencedirect.com/science/article/pii/S2949916X24000525?via%3Dihub

3. Key Benefits and Risks of Leveraging LLMs (Chatbot-based therapy) in the field of Mental Healthcare

Pros:

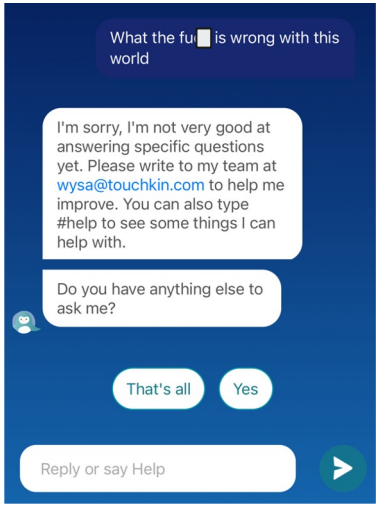
- Application is easily accessible to anyone who owns a smartphone and does not require in-person services
- The assistance provided and responses are generally unbiased
- Promotion of self-assurance & personal exploration at the users' pace



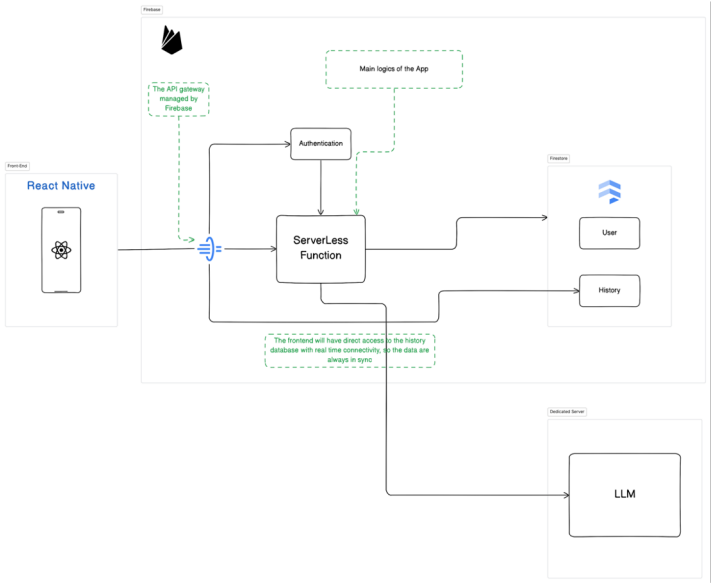


Cons:

- Lack of regulation of messages with profanity, harmful content and hate speech from other users
- Limited ability to maintain a consistent communication style
- Inability to retain long term memory and produce relevant, contextual responses catered to specific user experience
- User with censored messages does not get the relevant help needed



Our Solution of Addressing the Gap of Leveraging LLMs in the field of Mental Healthcare



A native chatbot app with a sandwich architecture that acts as a **moderator** between **user-user-LLM.**

2. Preliminary Results & Project Status

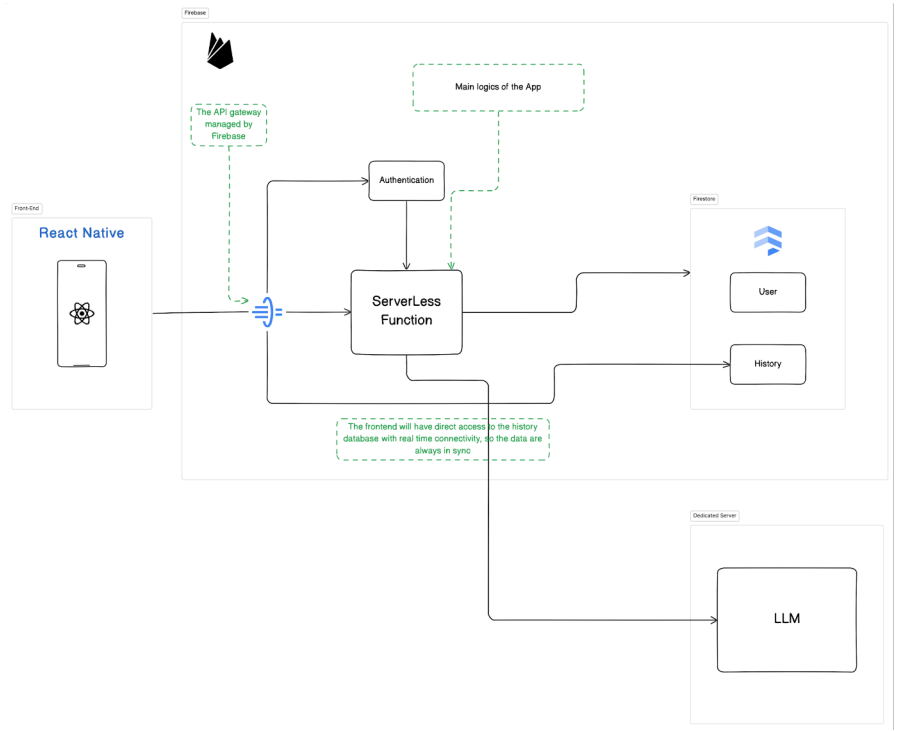
a. Solution to Address Identified Research Gaps- an architectural approach

a. Project Development Status
- Frontend
- Backend

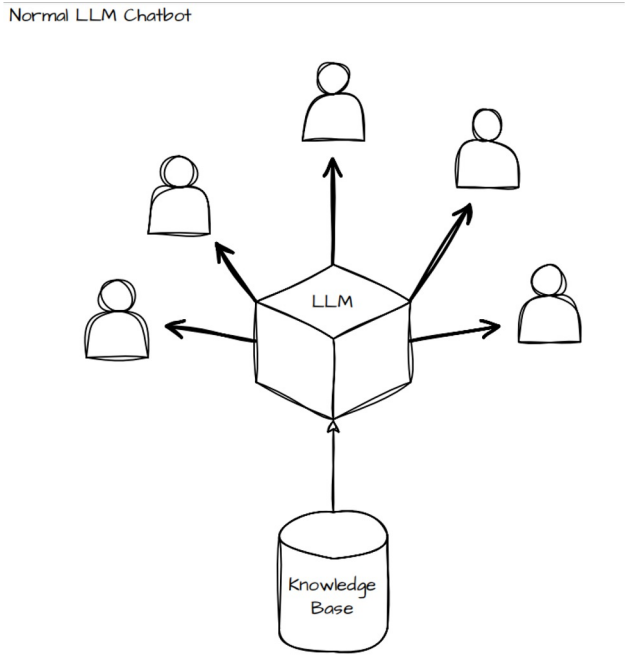
Sandwich Architecture

The sandwich architecture acts as a **moderator** between **user-user-LLM** to help address problems identified:

1. Content regulation
2. Communication consistency
3. Enhancement in empathetic & compassionate conversation
4. Reduction of bias & misinformation



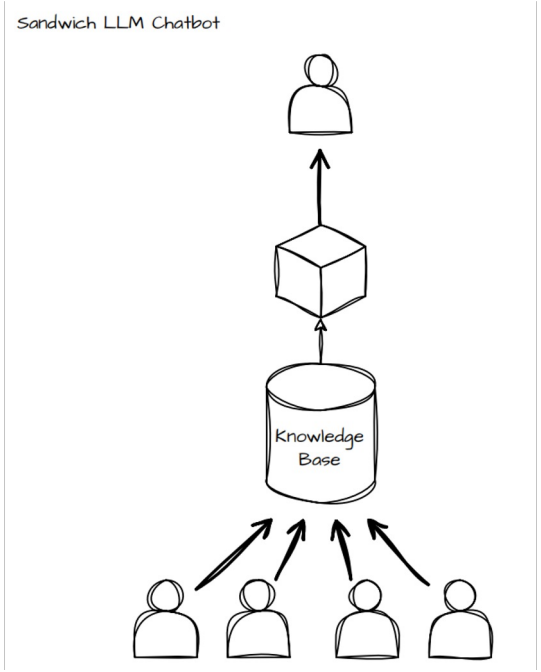
Main Idea of the Sandwich Architecture



All users share a **static** knowledge base

= Receiving Rigid & Standardized responses

Main Idea of the Sandwich Architecture



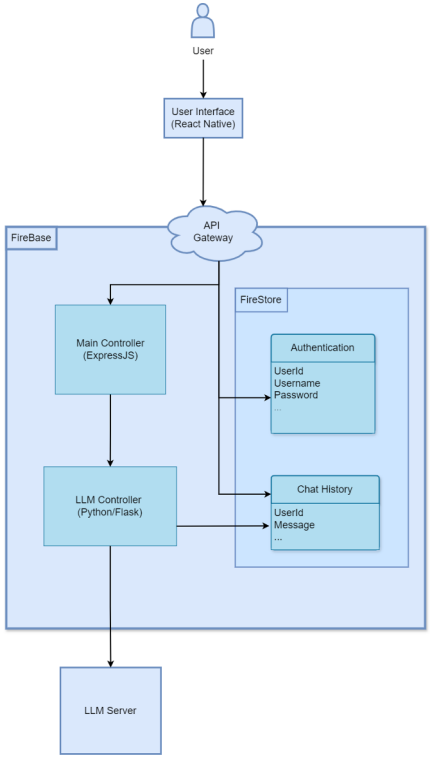
All users share a **dynamic** knowledge base

- Users receive responses from the LLM and contributes to the LLM of the knowledge base

- Users within a cluster interacts with the LLM, while their chat history is incorporated into the knowledge base, transforming them into both producers and receivers of information, thereby enhancing the flexibility and dynamism of the system

Main Idea of the Sandwich Architecture

System Design of Our Chatbot App



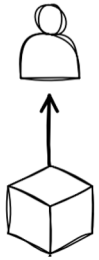
- System Design Overview:
  - User interface
  - Control logic
  - LLM server
- Major Access Control:
  - Control logic (back-end) deployed on Firebase, accessible only through API gateway
- Security solution:
  - Firestore utilized for databases
  - Provides integrated user authentication and data access
- System performance:
  - LLM model deployed on dedicated high-performance server
  - Only accepts requests from our system on Firebase

Main Idea of the Sandwich Architecture

Real-time User Interaction- Clustering

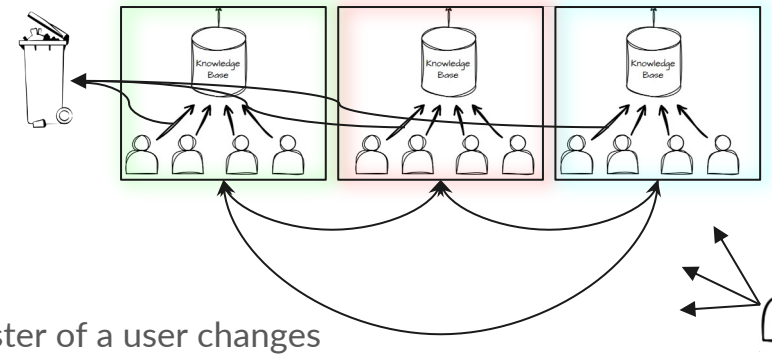
Sandwich LLM Chatbot

5) Chat records expire and are removed from time to time to prevent the cluster knowledge base from growing too large to maintain a real-time nature of the system.



5) Generation of contextual response to the user corresponding to the specific cluster by the LLM.

LLM



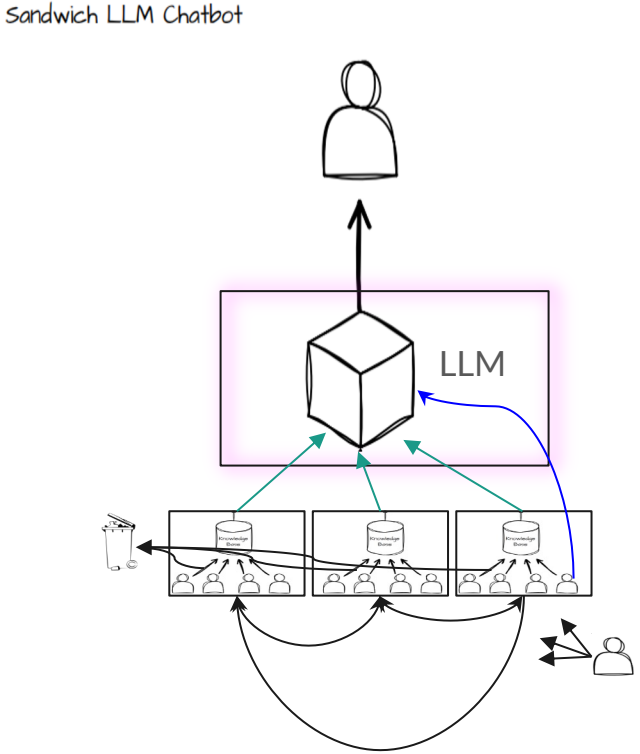
2) Each cluster maintains a specific cluster knowledge base and contains a number of users with similar behaviour and/or experience. i.e., the experience of isolation and loneliness during a mandatory quarantine imposed due to the COVID-19 pandemic's travel restrictions.

3) User chat history is integrated into the cluster knowledge base for the generation of contextual response from LLM.

4) Cluster of a user changes according to their behaviour.

1) New user is assigned to a cluster at random in the beginning.

Contextual Response Generation- Prompt Engineering

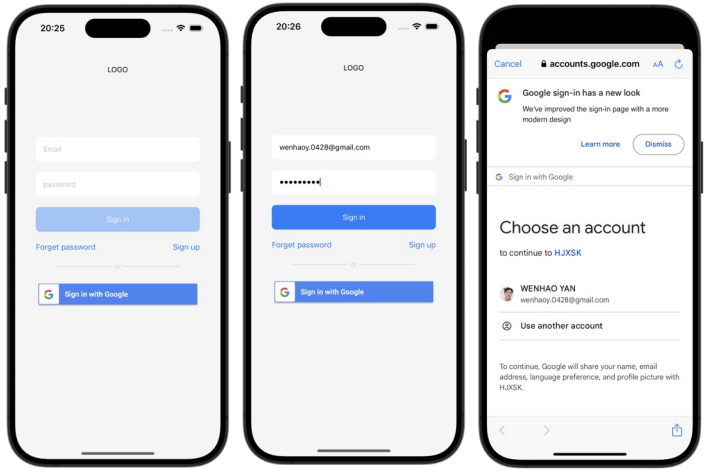


Enhance the specificity and sensitivity of LLM's responses by incorporating an **implicit prompt template:**

- Utilizes and extracts BOTH user's input (chat history), and the cluster knowledge base which has other users' chat history
- Inserts extracted details into the prompt template
- Improves the overall quality and contextuality of LLM's responses; and ensures the presence of a relevant response to the live user

Project Development Status (Preliminary Result)- Frontend

Major components: Sign-in, Sign-up, Chat Screen

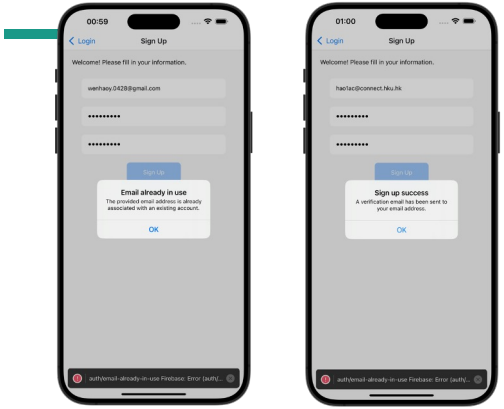


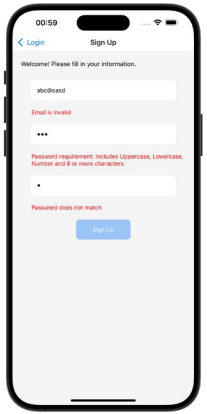
Key Features of the Sign-in Screen

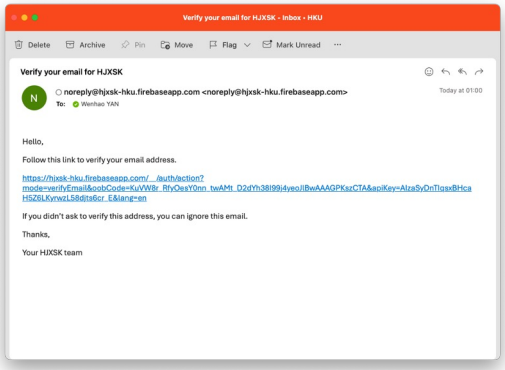
1. **Login Options**: The Sign-In Screen provides users with multiple options to authenticate their identities. They can choose to sign in using their registered Email/Password combination or opt for the convenience of Google Authentication. This flexibility caters to user preferences and enhances accessibility.
2. **Input Validation and Interactivity:** The screen incorporates intelligent input validation to ensure data accuracy and security. The Sign-In button remains disabled until all required fields are filled correctly, preventing users from submitting incomplete or erroneous information. This interactive design element guides users and fosters a sense of trust and confidence in the application.
3. **Streamlined User Experience**: The Sign-In Screen is designed to provide a seamless and efficient user experience. Clear and intuitive interface elements guide users through the sign-in process, minimizing any potential friction or confusion. The focus on usability allows users to quickly and effortlessly access their accounts and proceed to the application's features.

Project Development Status (Preliminary Result)- Frontend

Major components: Sign-in, Sign-up, Chat Screen







Key Features of the Sign-up Screen

1. **User Account Creation:** The sign-up form enables users to input their email address and password, ensuring the seamless creation of their account. This step is crucial for establishing personalized user experiences and ensuring secure access to the application's features.
2. **Error Handling and Alert Messages:** The Sign-Up Screen incorporates an intelligent error handling mechanism. If any errors occur during the sign-up process, the screen promptly displays an alert message, providing users with clear and concise feedback. This helps users identify and address any issues they may encounter during the registration process.
3. **Success Message and Verification Email:** Upon successful sign-up, the Sign-Up Screen displays a success message, assuring users that their account creation was completed successfully. Additionally, an automated verification email is sent to the user's provided email address, ensuring the security and authenticity of their account.

Project Development Status (Preliminary Result)- Frontend

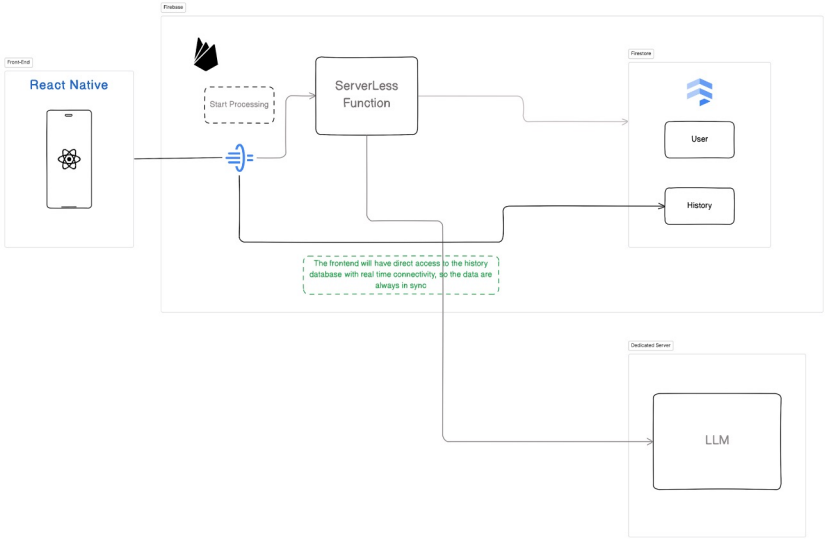
Major components: Sign-in, Sign-up, Chat Screen

Key Features of the Chat Screen

1. **Interactive Chat Interface**: The ChatScreenContent component presents an interactive chat interface that enables users to engage in a meaningful conversation with the chatbot. It provides a visually appealing and user-friendly design, fostering a natural and intuitive communication flow.
2. **User Input and Messages**: The chat interface allows users to input their messages, enabling them to interact with the chatbot effectively. It also displays the messages exchanged between the user and the chatbot, ensuring a clear and organized conversation history.
3. **Custom Message Bubble Component:** The ChatScreenContent component incorporates a custom message bubble component for rendering chat messages. This component enhances the visual appeal of the chat screen, distinguishing between user and chatbot messages, and promoting readability and comprehension.

Project Development Status (Preliminary Result)- Frontend

Major components: Sign-in, Sign-up, Chat Screen cont'



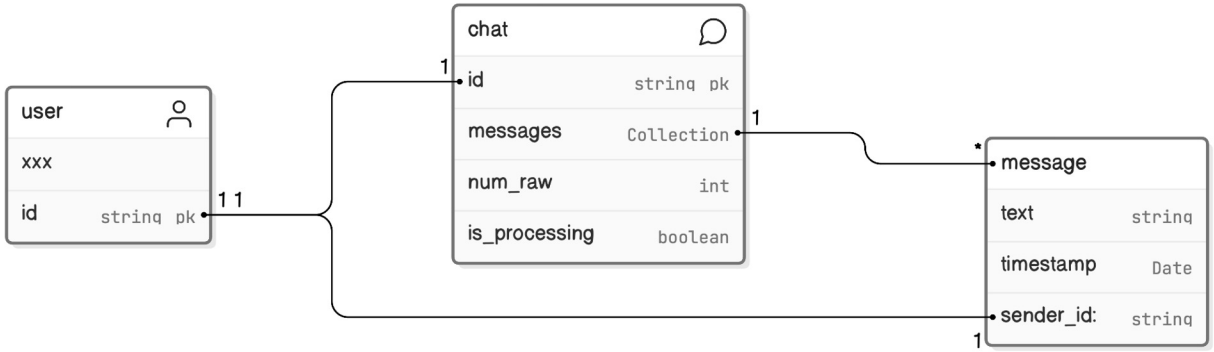
Optimization of the user experience considering the processing time and resource utilization of the Large Language Model (LLM). To balance the user expectations and system limitations, the frequency of interactions with the LLM is strategically reduced to maintain a smooth conversation experience for users.

1. **User Input Processing Optimization:** The chat screen employs a mechanism to send messages to the LLM only when the user has momentarily paused typing. This approach ensures that the LLM is triggered only when the user has completed composing a coherent message, enhancing the overall efficiency and accuracy of the LLM's responses.

1. **Backend Server Communication:** To indicate the start of processing, a server call is made to the backend when the user initiates a message. This communication serves as a signal to the server that the LLM processing is about to begin, allowing for effective resource allocation and optimal utilization.

Project Development Status (Preliminary Result)- Backend

Major components: Chat Schema



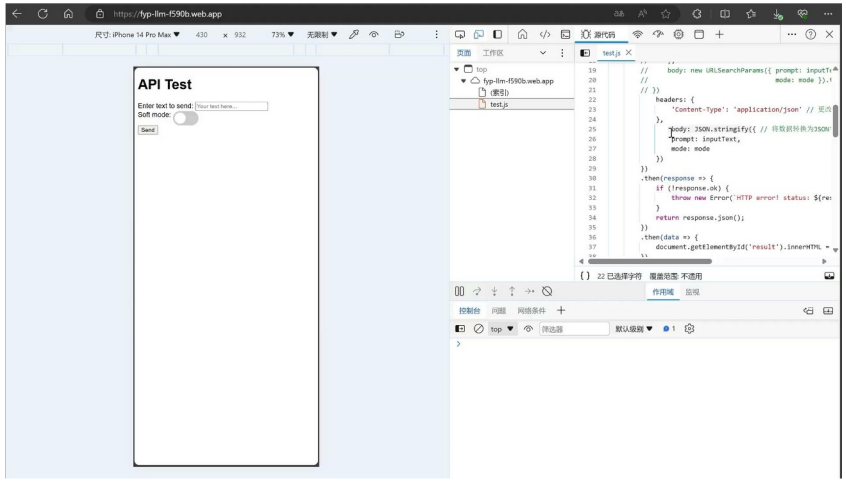
1. Persistent Storage & Real-Time Updates: The real-time capabilities of Firestore enable instant updates to the chat as new messages are sent and received, ensuring a fluid and dynamic conversation flow.

2. Bi-Directional Messaging: Both messages from user and LLM are promptly stored in Firestore's database in real-time. By capturing both messages, users who are subscribed to the database receive comprehensive and contextually relevant responses displayed on their screens.

3. Chat History Retrieval: The frontend of our application leverages Firestore's querying capabilities to retrieve the chat history from the database. By fetching the stored messages, the frontend can accurately display the conversation history to the user, providing a comprehensive view of the chat's progression.

Project Development Status (Preliminary Result)- Backend

Major components: Profanity Filter Demo



Identification and handling of bad intents within user input, transforming them into more gentle expressions. This implementation addresses two significant challenges:

1. **Mitigating Offensive Content Propagation:** The Profanity Filter acts as a safeguard, preventing the integration and dissemination of sentences with bad intents in the application. By detecting and filtering out offensive language, it ensures that such content is not sent to other users, thereby maintaining a respectful and inclusive environment for all.
2. **Enabling Response with Bad Intent:** Another challenge arises when the Large Language Model (LLM) encounters bad intent within user input, potentially leading to a refusal to provide a response. Our implementation overcomes this challenge by adjusting the way the LLM expresses its response. By transforming bad intent into more appropriate language, the Profanity Filter enables the LLM to generate a response that aligns with the application's desired tone and maintains a positive user experience.

API Test Reference Link: https://fyp-llm-f590b.firebaseapp.com/

3. Schedule for the Rest of the Project

Performing Comprehensive Debugging and Front-End/Back-End Integration	Start date: 1/June Soft deadline: 19/June Hard deadline: 26/June
Implementation of backend Clustering to complete the sandwich architecture design	Start date: 7/June Soft deadline: 14/June Hard deadline: 19/June
Implementation of Cybersecurity Measures: Inspecting User Messages and Encryption Enhancement	Start date: 7/June Soft deadline: 12/June Hard deadline: 21/June
Enhancing Settings/Personalization Front-End Experience	Start date: 7/June Soft deadline: 13/June Hard deadline: 20/June
Conduct Usability Study and send out questionnaire to assess the performance of the app in the context of being a mental health supporting chatbot	Start date: 26/June Soft deadline: 3/July Hard deadline: 5/July
Complete Project Website Link: https://hjxsk-hku.web.app/	Deadline: 8/ July
Complete Project Final Report	Deadline: 12/ July
Complete Project Presentation Powerpoint	Deadline: 15/ July

Question & Answer