Introduction:

In an era where mental health challenges are increasingly prevalent, the need for immediate, accessible support is crucial. With the growing demand for mental health services and shortages of mental health therapists, AI-powered solutions are emerging as a valuable tool for support, especially in situations where traditional therapy is not readily available. Our AI-powered mental health therapy bot with sentiment analysis bridges this gap by providing real-time personalized emotional support through engaging therapeutic conversations. The bot integrates cognitive behavioral therapy techniques, mindfulness exercises, and other resources to act as both a support tool for users and as a complementary resource for professionals in the mental health field. To ensure accurate, empathetic, and effective interactions, datasets, including conversational mental health data, sentiment recognition, therapist-patient interactions, and speech-based emotion recognition will help train the AI model to identify signs of stress, anxiety, and other emotional states.

Group Organization:

Group members and responsibilities.

Timothy Chan - Software Architecture, IT Infrastructure. - System Architect

Feiyang Li - Data Model, App Prototype (Therapist Side). - Information Officer

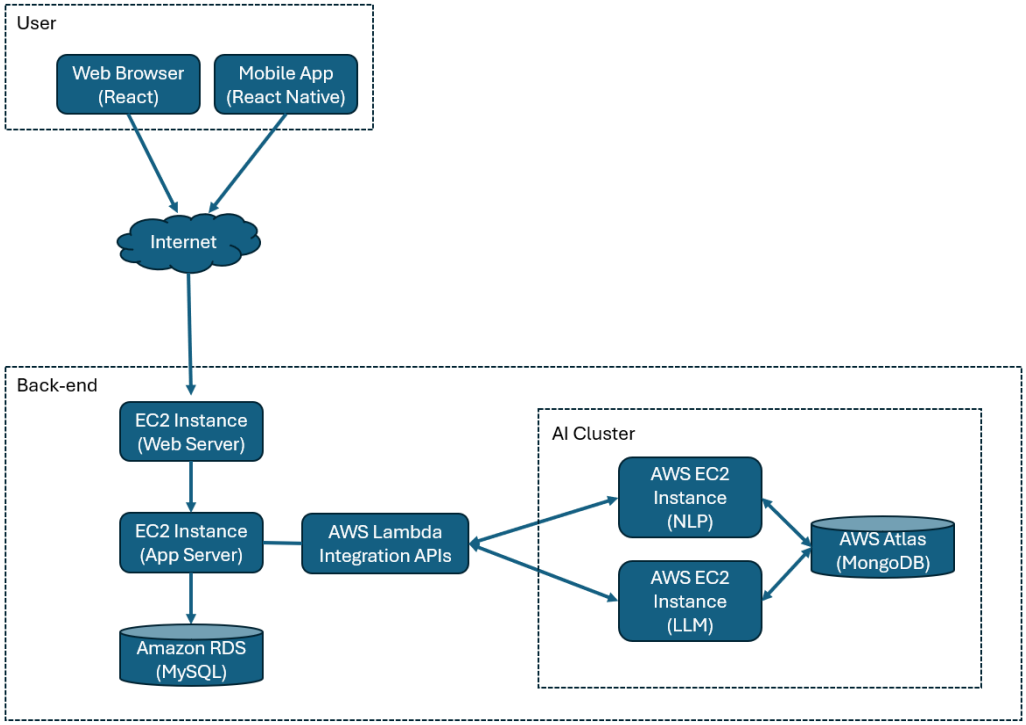
Vraj Patel - Business Aspects (Scalability), IT Infrastructure. - Information Officer

Eric Qiu - Functionality of the App, External Services. - Communication Officer

Daisy Ye - Business Aspects (Business Models), App Prototype (Client Side). - Project Manager

We are all in charge of putting the report and the slides together. For a more comprehensive work breakdown, please refer to Appendix A.

Software Architecture:



Data Model: No change.

IT Infrastructure List:

The infrastructure of our AI-powered mental health therapy bot is crucial for ensuring the system's scalability, reliability, and performance. The infrastructure components are detailed below:

- **Platform**:
    - **Cloud Provider**: AWS (Amazon Web Services) - EC2 Instances (t3.medium) for web and app servers. It provides a balanced mix of compute, memory, and network resources at a reasonable cost.
    - **Database**: MySQL for relational data and MongoDB for AI model weights.
    - **Load Balancer**: AWS Elastic Load Balancer (ELB) for distributing traffic across multiple servers.
    - **NVIDIA DGX H100** – AI training and inference server (6x H100 GPUs, 480GB total GPU memory)

    - **Dell PowerEdge R760 Server** – Compute and storage rack server (Intel Xeon Gold 6430 CPU, 64GB RAM, a 480 GB SSD).
- **App**:
    - **Web Application**: ReactJS for front-end development.
    - **Mobile Application**: React Native for cross-platform mobile development.
    - **Backend Framework**: Node.js with Express for API management.
- **Network**:
    - **Firewall**: AWS Security Groups to protect instances and ensure secure communication.
    - **VPN**: VPN Gateway for secure internal communication between services.
- **Data**:
    - **MySQL Database**: For structured data storage such as user information, session details, and therapy techniques.
    - **MongoDB**: For storing non-relational datasets, such as the AI model's weight and training data.

As for the platform and tools that were used to build the working prototype, we used Figma to design the mockup.

Services Offered to External Systems:

Our AI-powered mental therapy bot offers several essential services for external systems to better analyze and assess the mental well-being of individuals. One of the core services provided is the Emotional Support API. This API allows third-party applications, such as health apps and websites, employee well-being platforms in companies, and mediation and relaxation service providers to directly use this API to enable AI-driven therapeutic conversations that improve their services' quality and performance. This API can analyze the user's sentiment and provide real-time analysis and interactions. When used in the external systems mentioned, it can act as a second choice for users who can not get an appointment with a therapist or are in desperate need to resolve their anxiousness. From a technical point of view, our API can be accessible through HTTPS, and to utilize this service, external systems need to send a POST request to our secure endpoint at https://api.mentaltherapybot.com/v1/emotional-support,

including an API key in the header for us to correctly authenticate them. The request body must contain the exact data in JSON format following our requirements, which includes a user-id, conversation-id, prompt from the user, and a timestamp of when the user has input the prompt. After receiving the data, our system will check if all the data follows the correct format. For example, the parameters user-id must be valid containing only letters and numbers, and is within the character limit. Also, we will check if the user prompt is within the character limit, and if the timestamp is correctly formatted as required. In cases of detecting these errors mentioned, the API will communicate clearly through HTTPS error codes such as “500” and “400”, followed by the corresponding output explaining what the error is in text format.

Furthermore, Our system also provides an analysis and dashboard of each user’s mental health situation. With certain security and privacy measures in place, an authorized therapist or doctor can gain direct access to the patient’s data stored in our database, and be able to view detailed analysis through interactive dashboards provided so that they can make more informed diagnoses and provide better decisions to resolve the patient’s problems. To gain access to this service, we offer connection through an HTTPS link https://dashboard.mentaltherapybot.com.

For each authorized therapist or doctor, they will have their unique login credentials created and approved by our team, and each time they want to use this service, they must use this login information and go through a 2-factor authentication for security purposes. In addition, our security measures include SSL encryption, which encrypts the data transmitted between the user's browser and the server, input validation, and secure credential storage to provide top-level security to our customer’s information. To handle errors such as invalid login information or session timeout, the system will output a corresponding user-friendly message to the interface to communicate effectively.

Cost Estimation:

Here we will estimate the cost of building the AI model as well as the cost of maintaining the front-end and back-end server, giving the infrastructure that we decided in the previous section.

Estimation of the cloud infrastructure (AWS)

1. EC2 Instances (t3.medium)
    a. Estimated cost of $0.0416/hour
    b. We decided on running with 3 instances 24/7, because we want our servers to run day and night and 3 instances ensure that our servers are less likely to go down.
    c. Cost estimated: 3 * 24 * 30 * 0.0416 = ~$90/month
2. AWS Elastic Load Balancer (ELB)
    a. Load balancer cost: ~$0.0225 per LCU-hour
    b. Estimated ~$17/month
3. AWS Security Groups:
    a. No additional cost
4. VPN Gateway:
    a. AWS Site-to-Site VPN: ~$0.05/hour
    b. Estimated: ~$36/month
5. Database Storage (AWS RDS MySQL)
    a. AWS RDS MySQL instance (db.t3.medium): ~$0.067/hour
    b. Cost: ~$48/month
    c. Storage (200 GB SSD): ~$0.10/GB
    d. Cost: $20/month
6. MongoDB (AWS DocumentDB)
    a. Instance (db.t3.medium): ~$0.06/hour
    b. Cost: ~$45/month
    c. Storage (100 GB SSD): $10/month

AWS Service estimated cost of ~$266/month, this would be an on-going cost.

Estimation of the AI Training & Compute cost:

1. NVIDIA DGX H100 Server (6x H100 GPUs, 480GB total GPU memory)
    - Cost: ~$350,000 one-time
    - Power & cooling: ~$1,500/month
    - Maintenance & support: ~$1,000/month
2. Dell PowerEdge R760 Server

- Cost: ~$10,000 one-time
- Power & maintenance: ~$500/month

Therefore the one-time purchase cost of: ~$360,000 and a monthly operating cost of ~$3000/month, however this monthly cost would incur because we want to constantly improve our AI model.

Development of the software cost:

1. Web Application (ReactJS) & Backend (Node.js + Express)
    a. Developer cost: ~$50/hour
    b. Assume 2 developers for 4 months full-time (8 hours work day)
    c. Cost: 2 developers * 4 month * 30 days/month * 8 hours/day * $50 = $96,000
    d. Assumption that the developers hired are freelancers, no need for the office cost.
2. Mobile Application (React Native)
    a. Developer cost: ~$50/hour
    b. Assume 1 developer for 4 months
    c. Cost: 1 * 4 * 30 * 8 * 50 = $48,000
3. UI/UX Design (Figma)
    a. Designer cost: ~$40/hour
    b. Assume 1 designer for 1 month
    c. Cost: 1 × 1 * 30 * 8 * 40 = $9,600
4. QA Testing
    a. Tester cost: ~$40/hour
    b. Assume 1 tester for 2 months
    c. $9,600

Therefore estimated one-time cost, incurred within 4 months of: $163,200 for development of the software. However, after the development is complete we still need at least 1 developer to maintain the application therefore at minimum require a ~$12,000/months

The security and compliance cost:

1. SSL Certificate (for HTTPS API & Dashboard)

    a. Estimated $300/year
    b. $25/month
2. 2-Factor Authentication (Auth0, Okta, or AWS Cognito)
    a. Estimated $500/month
3. Secure Credential Storage (AWS Secrets Manager)
    a. Estimated $40/month
4. Legal & Compliance (HIPAA, GDPR)
    a. Estimated $10,000 for audits

Hence, we estimated a one-time cost of $10,000, and $565/month.

Lastly, we created some API this require API hosting cost:

1. Assume 1,000,000 API calls/month
2. AWS Lambda or EC2-based API gateway: ~$0.02 per 1,000 requests
3. Cost: ~$20/month

And the dashboard hosting is covered under the AWS database. Therefore total cost of $20/month (on-going)

Also we would hire some therapist support team, and assume

1. initially we only hire 2 therapists
2. $40 per hour (4 hours per week)
3. Monthly cost of: 2 people * 30 days * 8 hours * $40/hour = $19, 200

So the cost of the support therapist would be $19,200/ month

Based on the cost analysis we would have an on-going cost $35,051/month and a one-time cost of $542,800. The data mainly obtained from AWS main website, and the average Ontario wage for different occupations. In addition, this cost is estimated with a minimalist assumption, the cost only based on the cost of the software system; however, in real life there is usually cost involved with marketing, finance, sale, taxation, and organization.

Our first years of the operation cost would be around $962 thousands (35 * 12 + 542), approximately $1 million. To investors we would ask for around $2 millions for the first year total cost which includes both 1 million operation cost and 1 million of other costs.

Conclusion:

In conclusion, our AI-powered mental health AI bot addressed the need of the accessible and scalable mental health support, by combining the ai technology and therapeutic methodologies our system provides real-time emotional assistance while ensuring support to healthcare professionals by providing data-driven insights. This is achieved by our robust architecture in combination with sturdy infrastructure design to ensure a high performance and strong security system. Lastly, indicated by our cost estimation we would require a cost of $542 thousands to construct the software, and $35 thousands monthly maintenance cost; in combination with marketing and operation cost, $2 millions required for this business to thrive in the first year.

Appendix A - Project Diary:

Date: 2025/03/14

Milestone: Second Phase of Project

Progress Overview:

- Completed the CV, introduction, group organization, architecture, and data model.
- Assigned main components and deliverables:
    - Leads: Daisy (partial), Timothy
    - Assumed to involve physical and system components necessary for deployment.
  - TCO - Total Cost Estimation (20 pts):
    - Leads: Feiyang, Vraj
    - Will depend on finalized infrastructure specs.
  - Offering Services (15 pts):
    - Leads: Eric, Vraj
    - Focused on defining what services the platform will provide and how.
  - Conclusion (2 pts) and Appendix (5 pts):
    - To be completed at the end of the documentation phase.

Demo:

- Team Involved: Daisy, Feiyang
- Demo Format: Eric has outlined preliminary ideas for the format.
- Next steps include refining flow and assigning specific roles for presentation/delivery.

Timeline Goal:

- Aim to finalize report by: Sunday, March 16, 2025

Next Meeting:

- Date: Sunday, March 16, 2025
- Time: 9:00 PM
- Topic: Report review & demo prep

Date: 2025/03/16

Milestone: Development Process Check-in

Feature Breakdown (Platform Process):

1. Home & Menu/Login Page – Completed
2. Chatbot Pages:
    - One for generating AI prompts
    - One for providing tips post-selection
3. FaceTime with Therapist – In Progress
4. Therapist View (Dashboard):
    - Interactive interface displaying patient data
5. UX/UI of Prototype:
    - Clarification Needed – likely refers to polishing front-end elements, layout consistency, and user experience design for the overall prototype.

Appendix B - Reason in regard to doing the demo using video instead of presenting it in class.

We decided that it was better to present using a premade video because both Daisy and I did a few test runs on Sunday and were not confident in doing it during the class. On the first few test runs, we were out of the time limit, and in the later run, there was a lot of stuttering. Later that night when we went through the guidelines we saw the sentence:

"Phase III can include making a video (for bonus marks) to demonstrate the final version of your system, which you can use to showcase the working integrations."

So we thought that a pre-made video would be acceptable; hence we did that.