AI Mental Health Application Testing

CMPE 187
Dr. Jerry Gao
By: Ngan Luu, Tiffany Nguyen, Prabjyot Obhi, Wenyi Cai
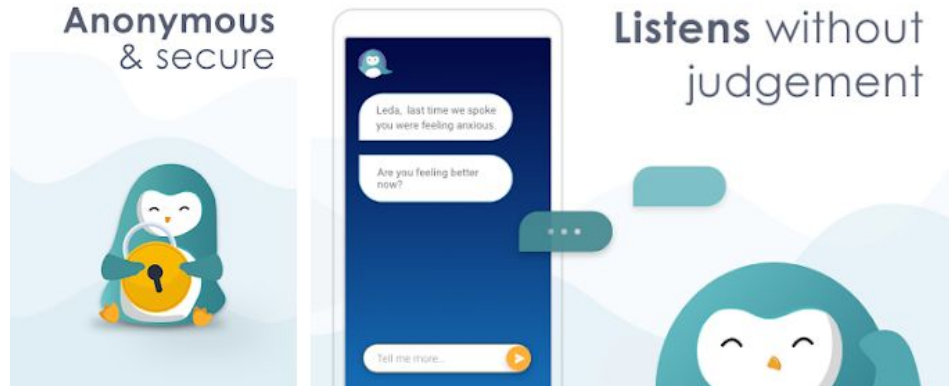
Application Overview

1) Artificial Intelligence (AI) Software that our team selected to test is Wysa mobile Application

2) Wysa is a machine learning based on AI chatbot system.

3) Wysa is a conversational AI that uses chatbot to interact with people in a humanlike way.

AI Chatbot (Wysa)

- The ability to stimulate a conversation with a user in English through messaging.
- In the context of app, the ability is to help the users, who are struggling with emotions or mental health, to share their feelings and help them feel better

Anonymous & secure

Leda, last time we spoke you were feeling anxious.

Are you feeling better now?

Tell me more..

Listens without judgement

How does Wysa work?

- The input required to use the application, is 2 types:
  - Selecting fast reply suggestion button
  - Type in answer/question
- The output is the conversation between Wysa and user

Our Objectives

1) Analyze software requirements
2) Define scenarios
3) Write and generate test cases
4) Find at least 1 bug within the resources we have
5) Review results
6) Validate
7) Use automation in testing

Conventional Test Methods

- Black Box Testing:
    - Decision table
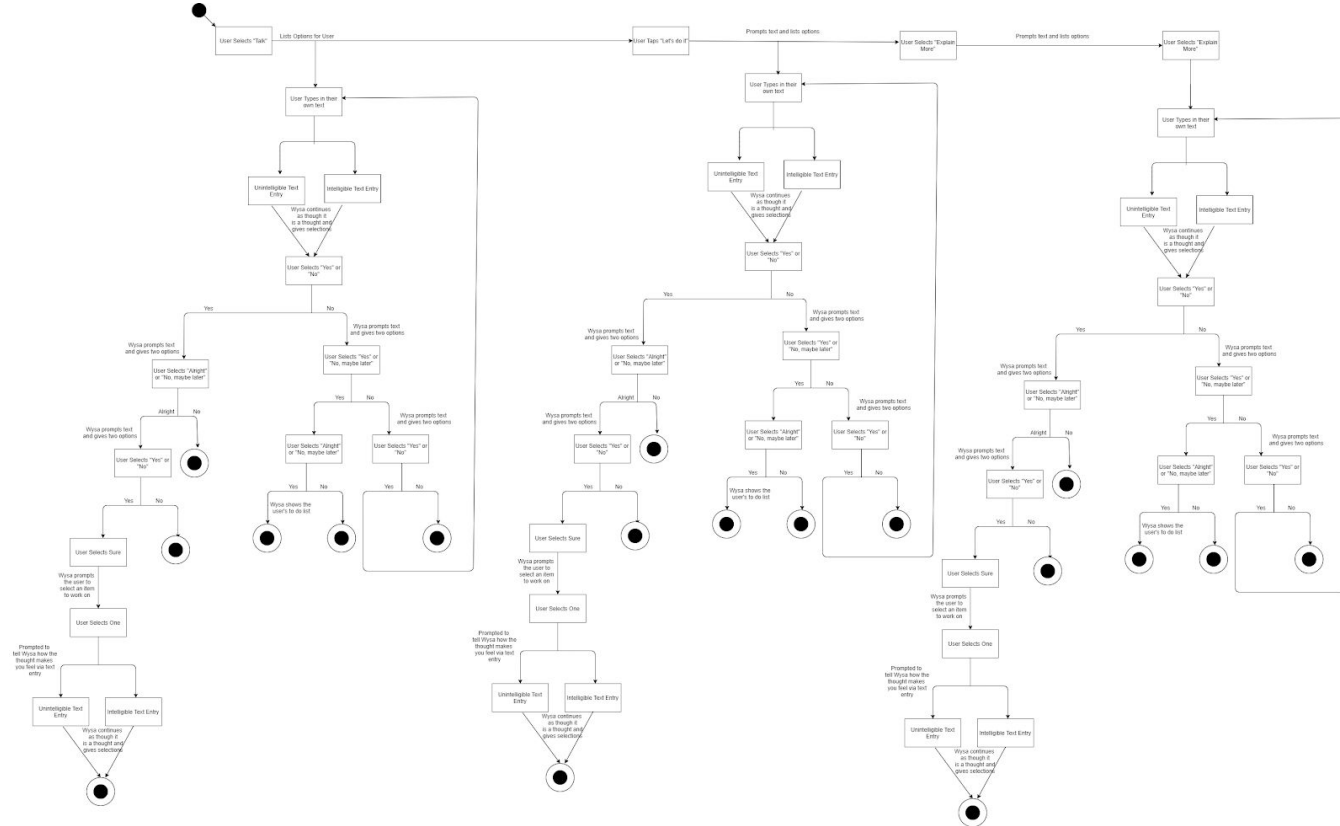   - Scenario testing
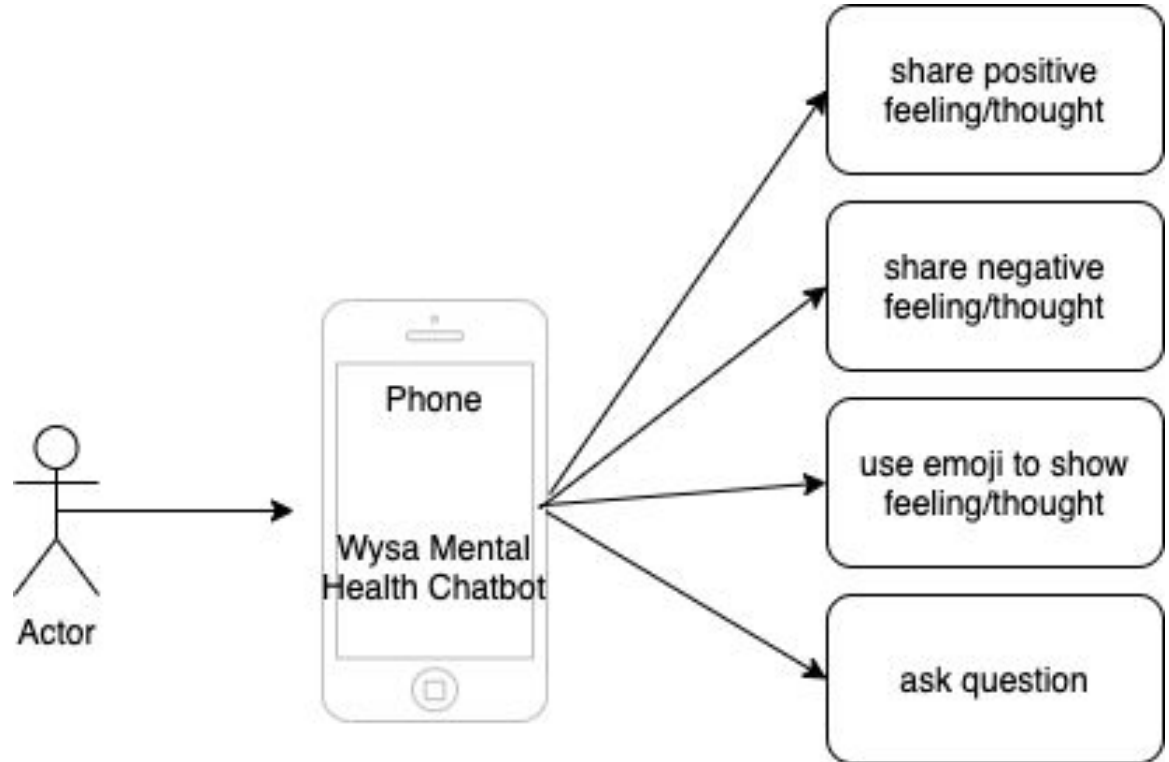
Scenario Testing

1. User has depression and wants an accessible way for therapy
2. User has anxiety and is looking for exercises through an accessible platform
3. User is self-deprecating and looking for ways to feel more confident (want to be in a happier mood)
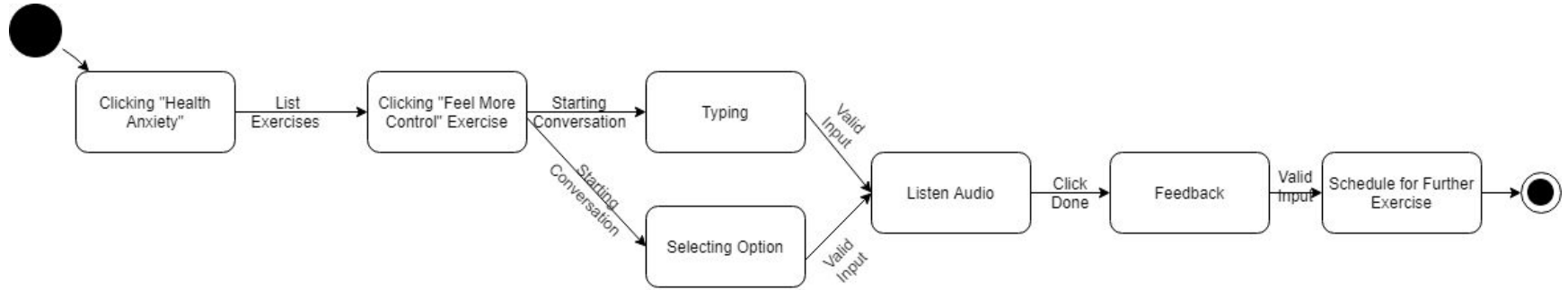4. User is struggling to sleep at night and looking for ways to ease themselves
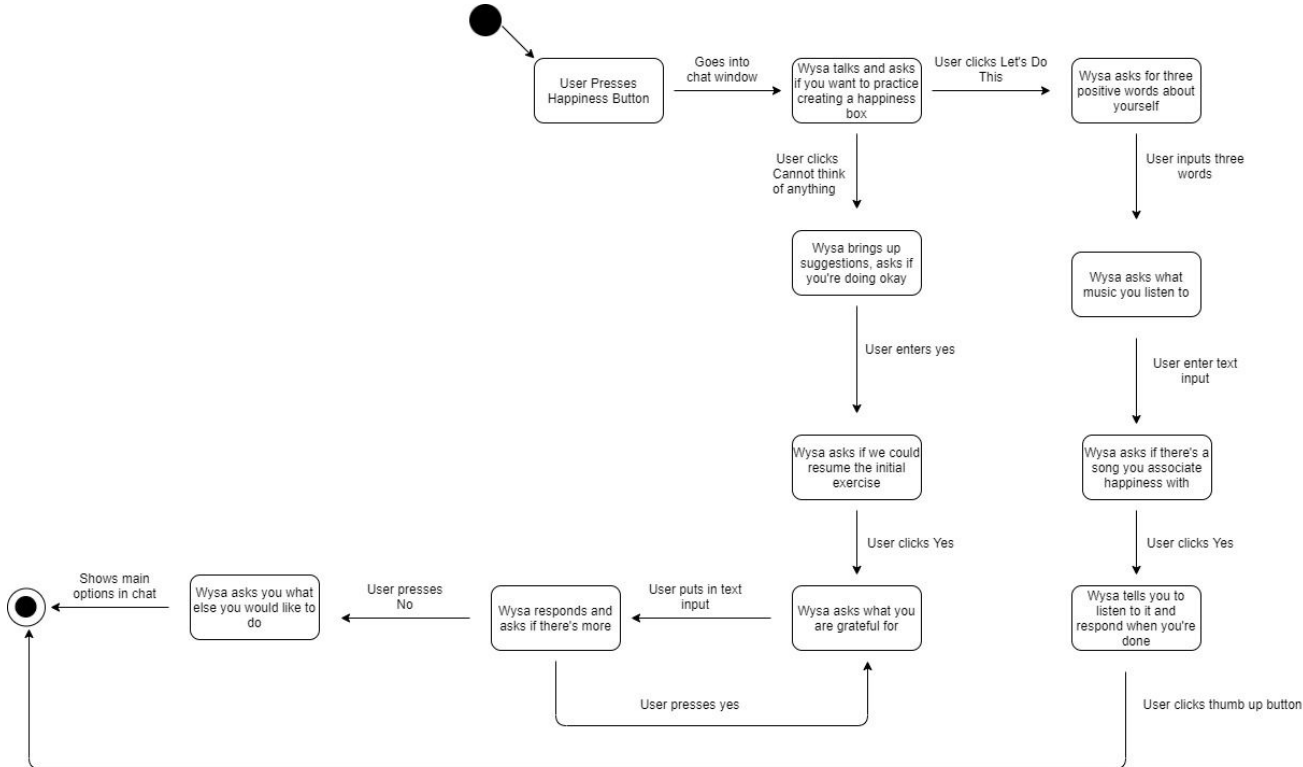
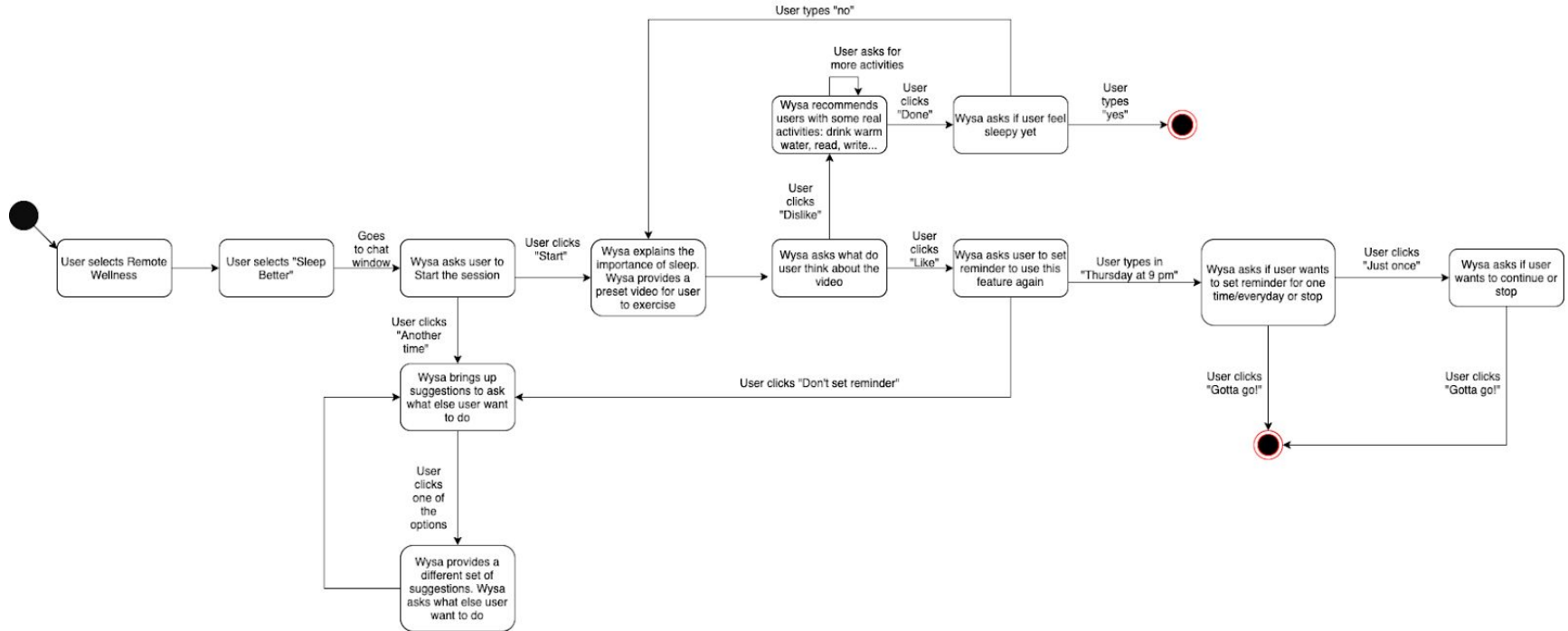# Scope for Testing Wysa
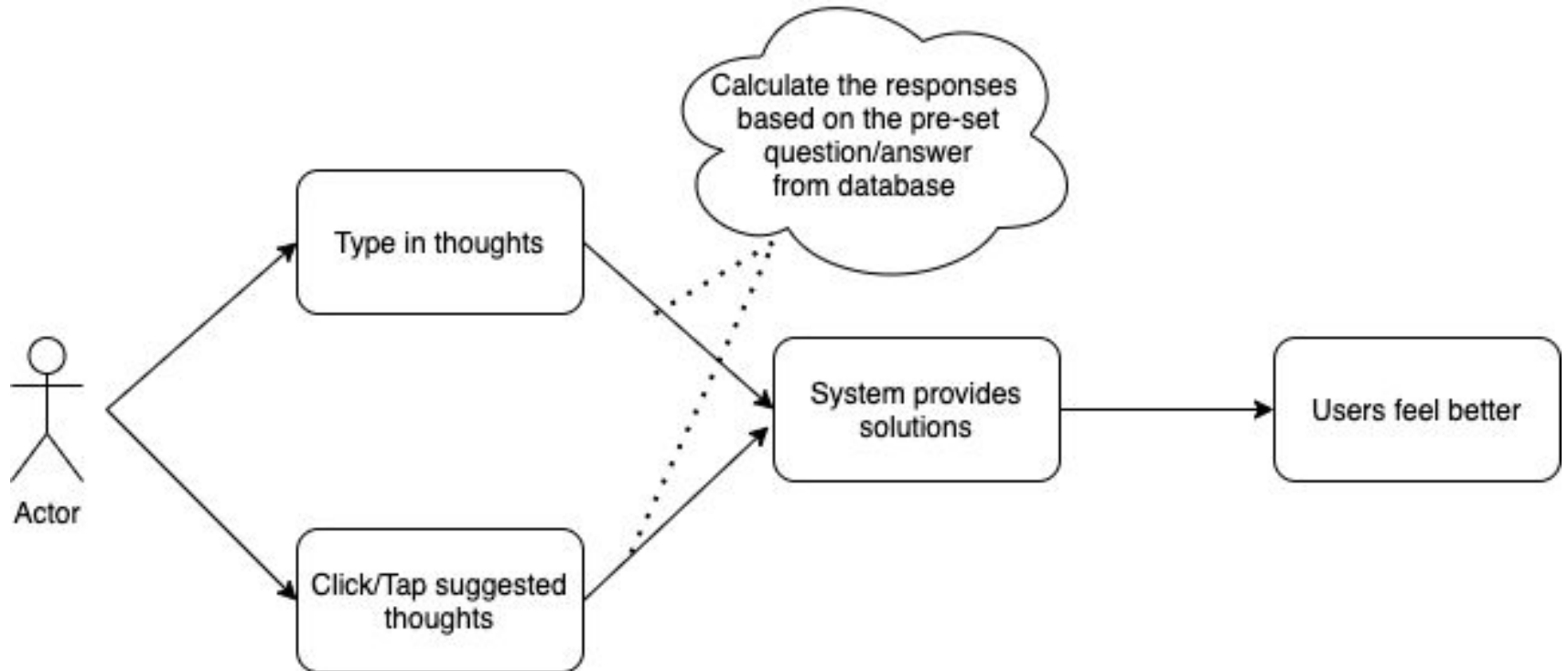
Diagrams from Scenario Testing (¼)

Diagrams from Scenario Testing (2/4)

Diagrams from Scenario Testing (3/4)

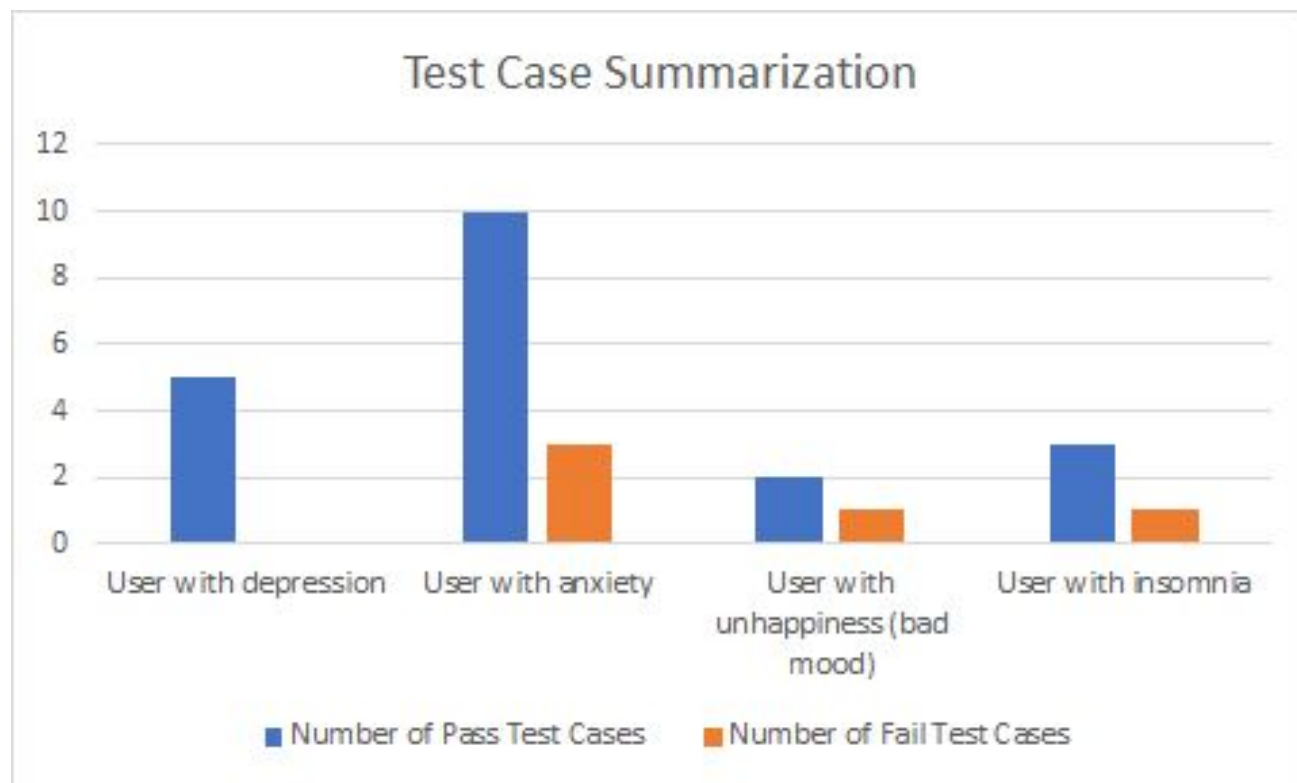# Diagrams from Scenario Testing (4/4)

Some Test Cases..

1. For someone who has depression, some cases:
   a. Needing someone to talk to who will comfort them
   b. Experiencing serious thoughts like suicide
2. For someone who is in a bad mood, some cases:
   a. Responded something sarcastically
   b. Still thinking pessimistically
3. For someone experiencing anxiety, some cases:
   a. Wanted to try a different exercise
   b. Was not satisfied with the exercise Wysa recommended
4. For someone who is experiencing insomnia, some cases:
   a. Wanted to do more than one exercise today
   b. Wants a different exercise from the day before

Results from Scenario Testing

Scenario	Description	Number of Pass Test Cases	Number of Fail Test Cases
1	User with depression	5	0
2	User with anxiety	10	3
3	User with unhappiness (bad mood)	2	1
4	User with insomnia	3	1

Results from Scenario Testing

Test Case Summarization

Bar chart comparing Number of Pass Test Cases (blue) and Number of Fail Test Cases (orange) across user scenarios:
- User with depression: Pass 5, Fail 0
- User with anxiety: Pass 10, Fail 3
- User with unhappiness (bad mood): Pass 2, Fail 1
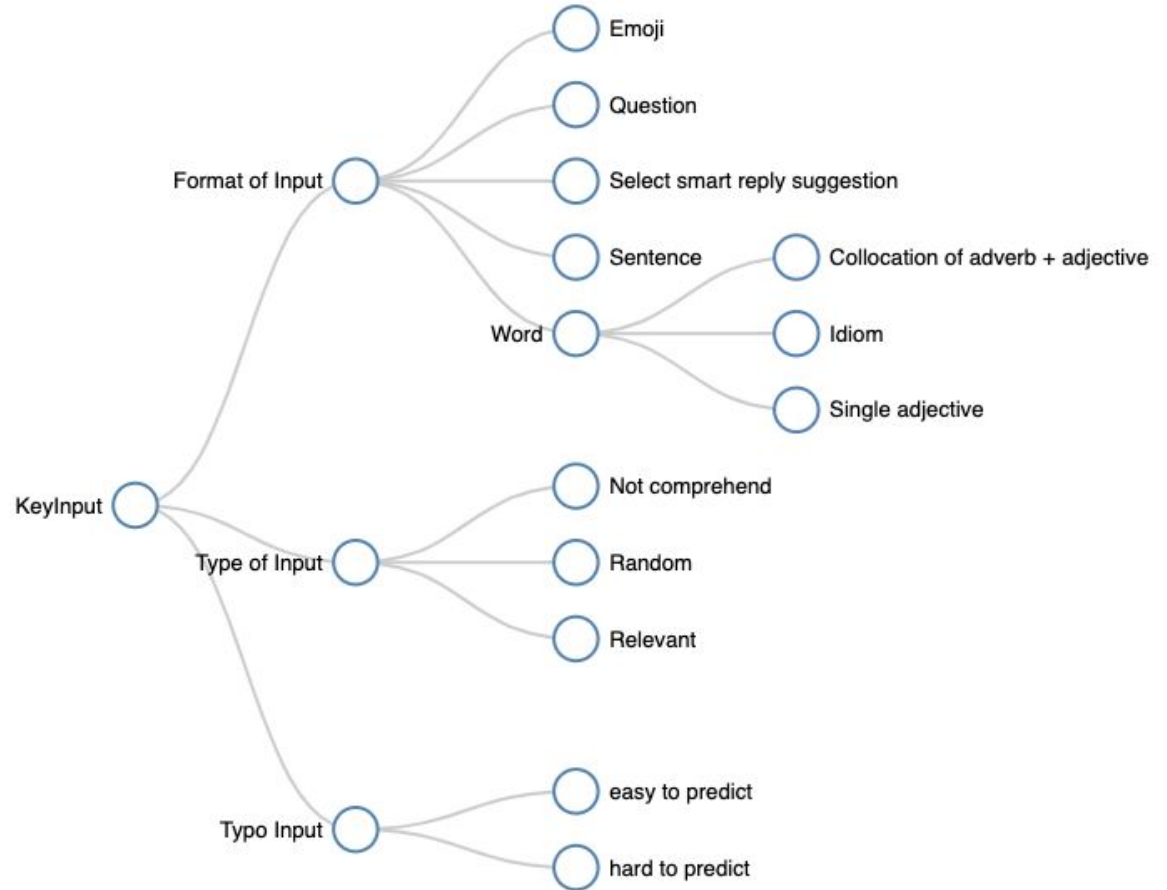- User with insomnia: Pass 3, Fail 1

Test Model - AI Based Tree - AI Chatbot

Context Classification Spanning Tree
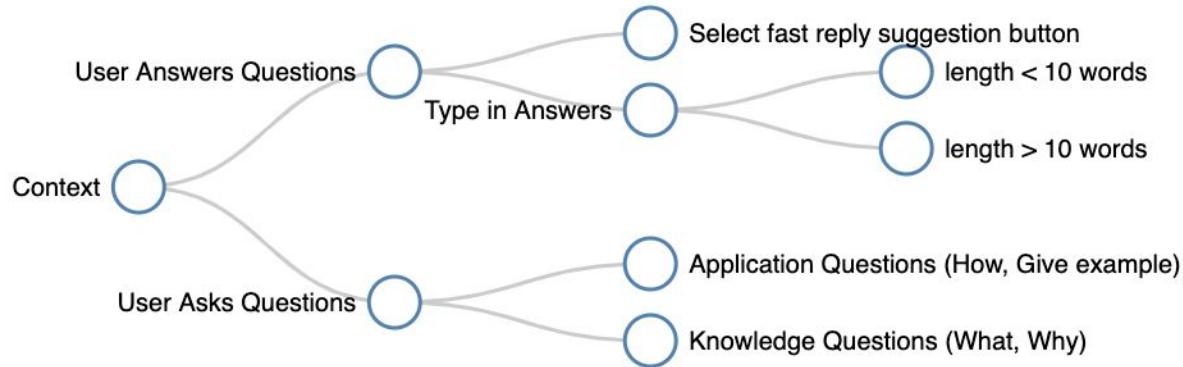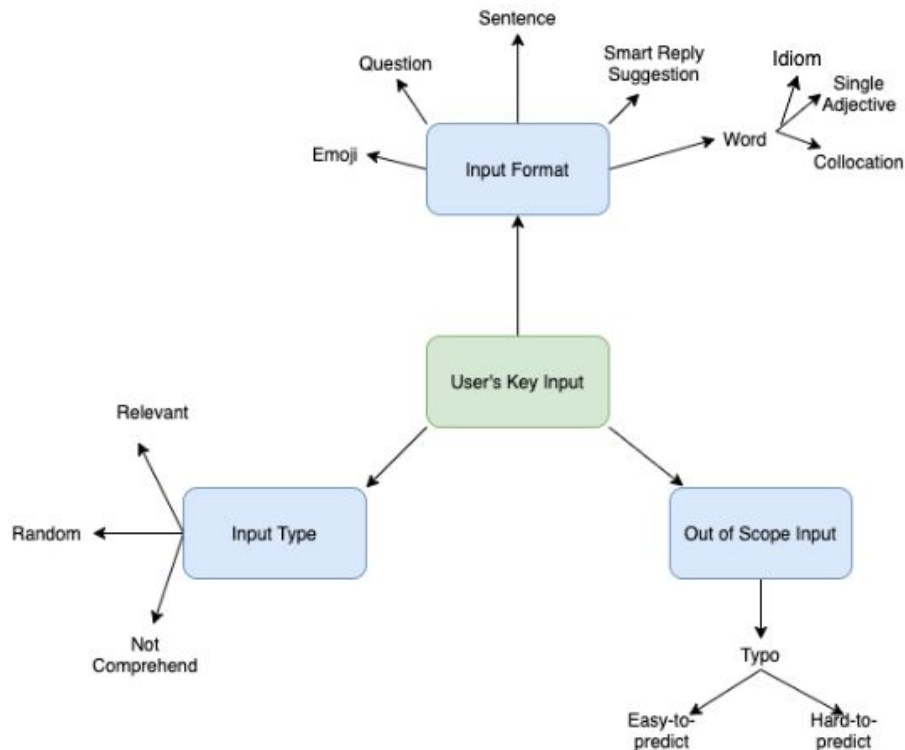
Test Model

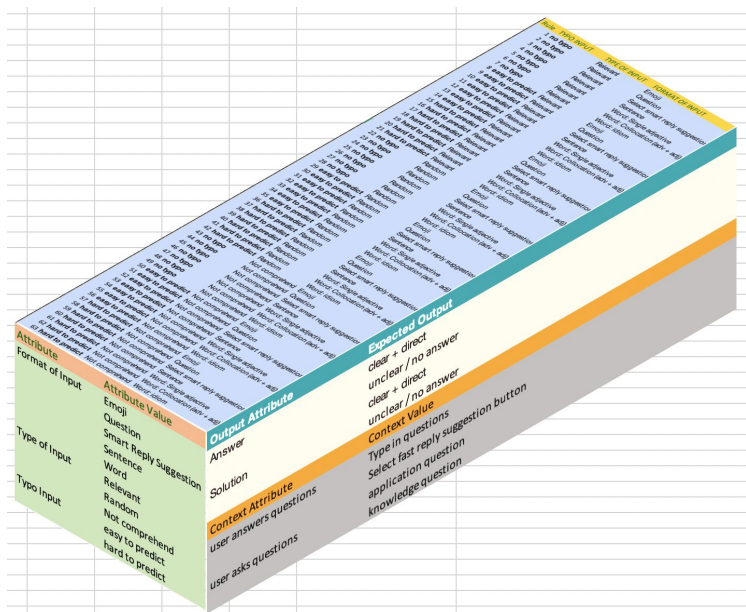Input Classification
Spanning Tree

# Output Classification Spanning Tree
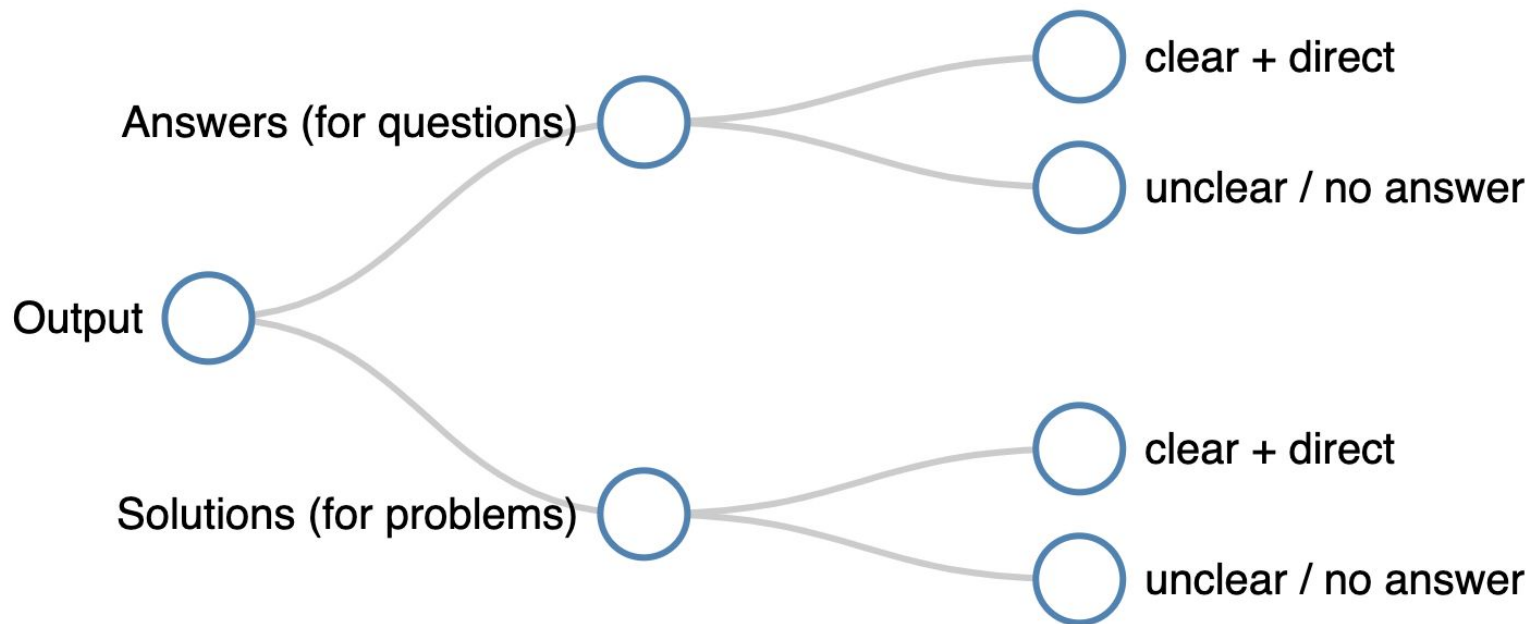
Test Data Model

We generated 18 test cases

Testing Cases

- Users use the chart box to type in answers with short length < 10 words
- Users use the chart box to type in answers with short length >= 10 words
- Users select answers through the reply suggestion button
- Users ask "What…."questions
- Users ask "Why…."questions
- Users ask application questions
- Users type Emoji
- Users use the smart reply button to chat.
- Users type "easy-predict" type words
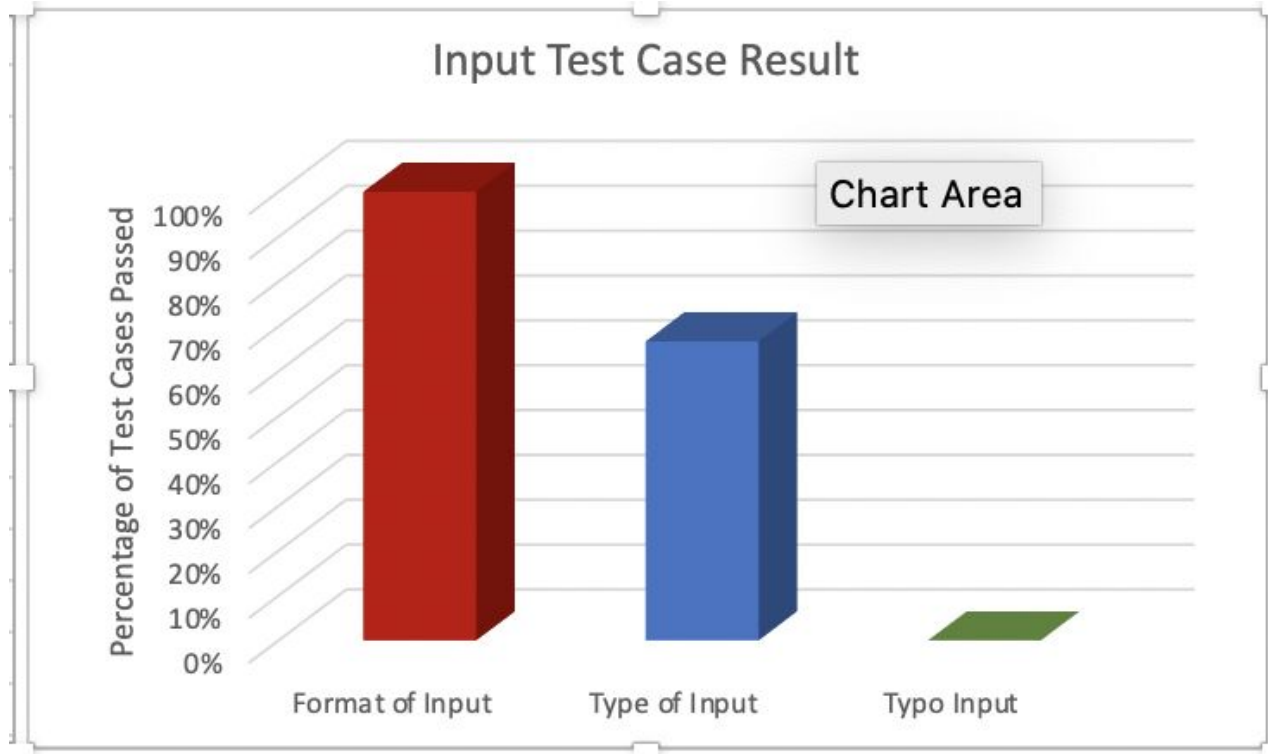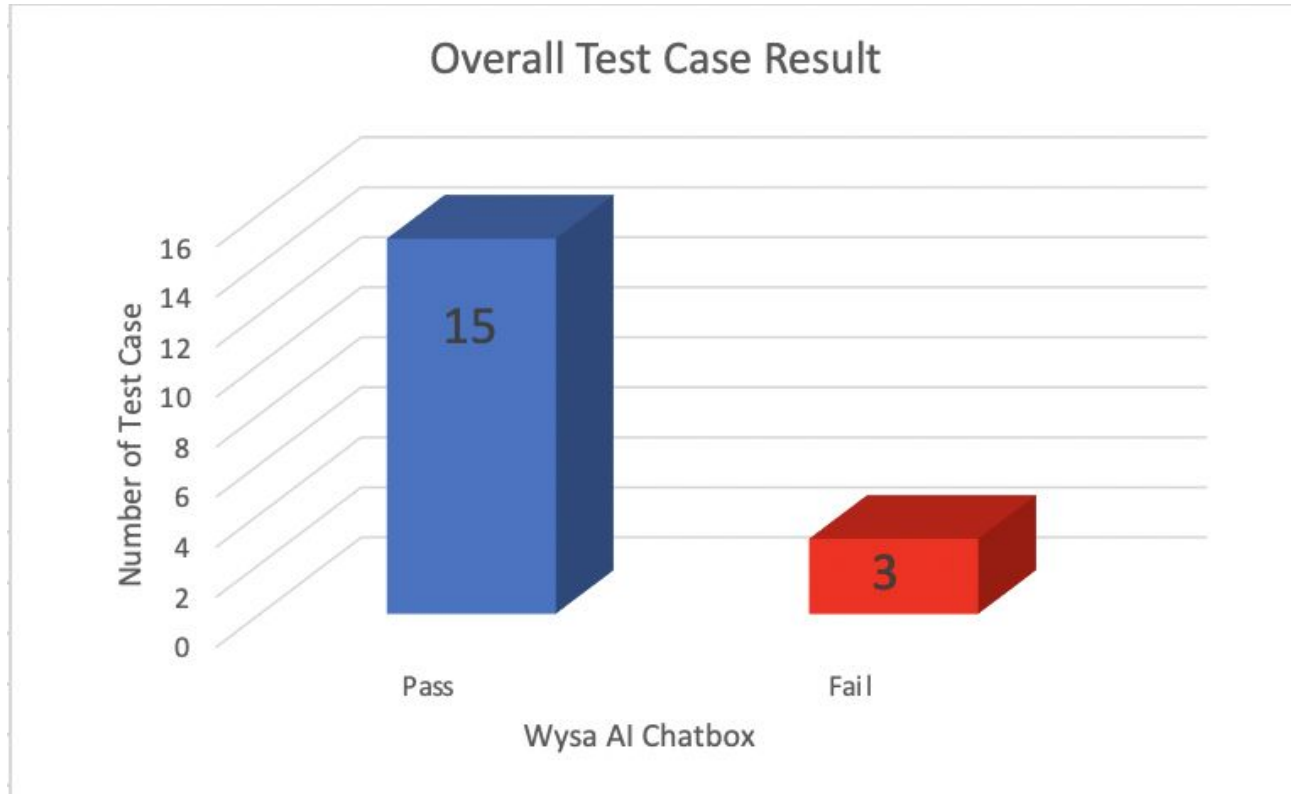- Users type "hard-predict" Type words.

Test Results
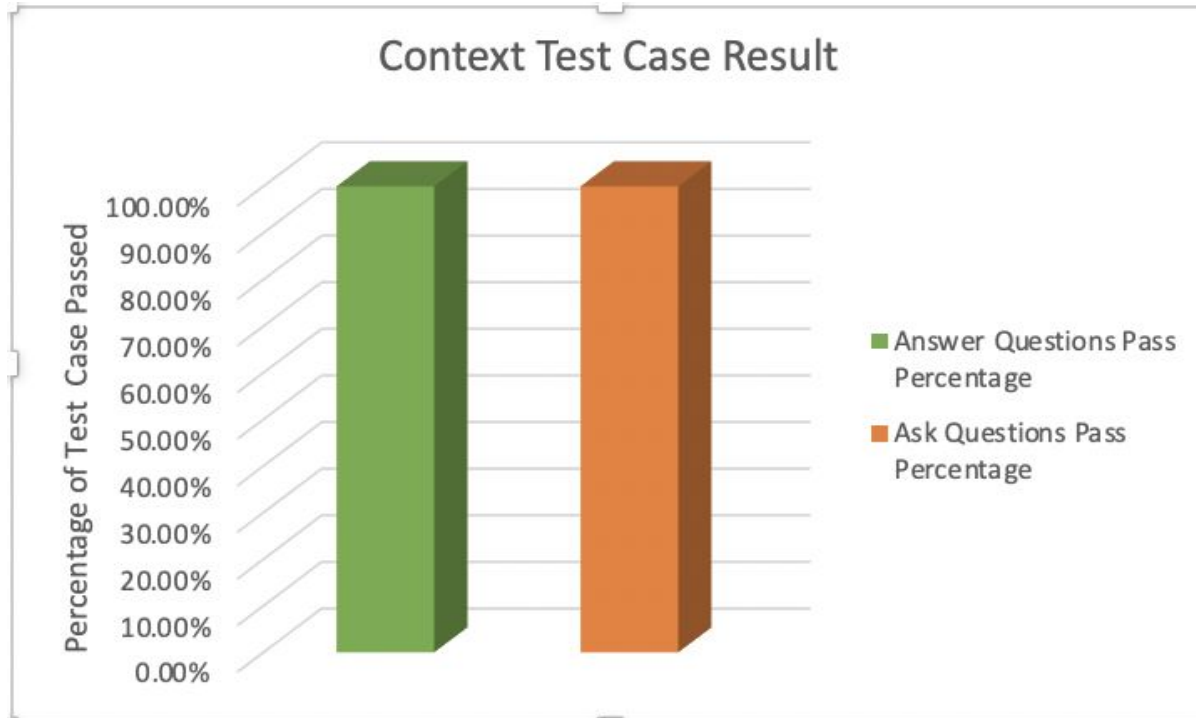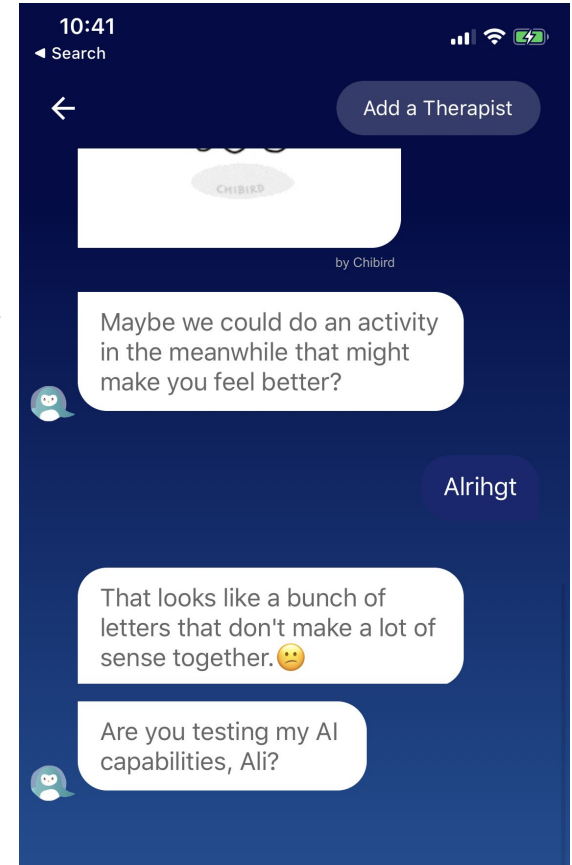
Test Results
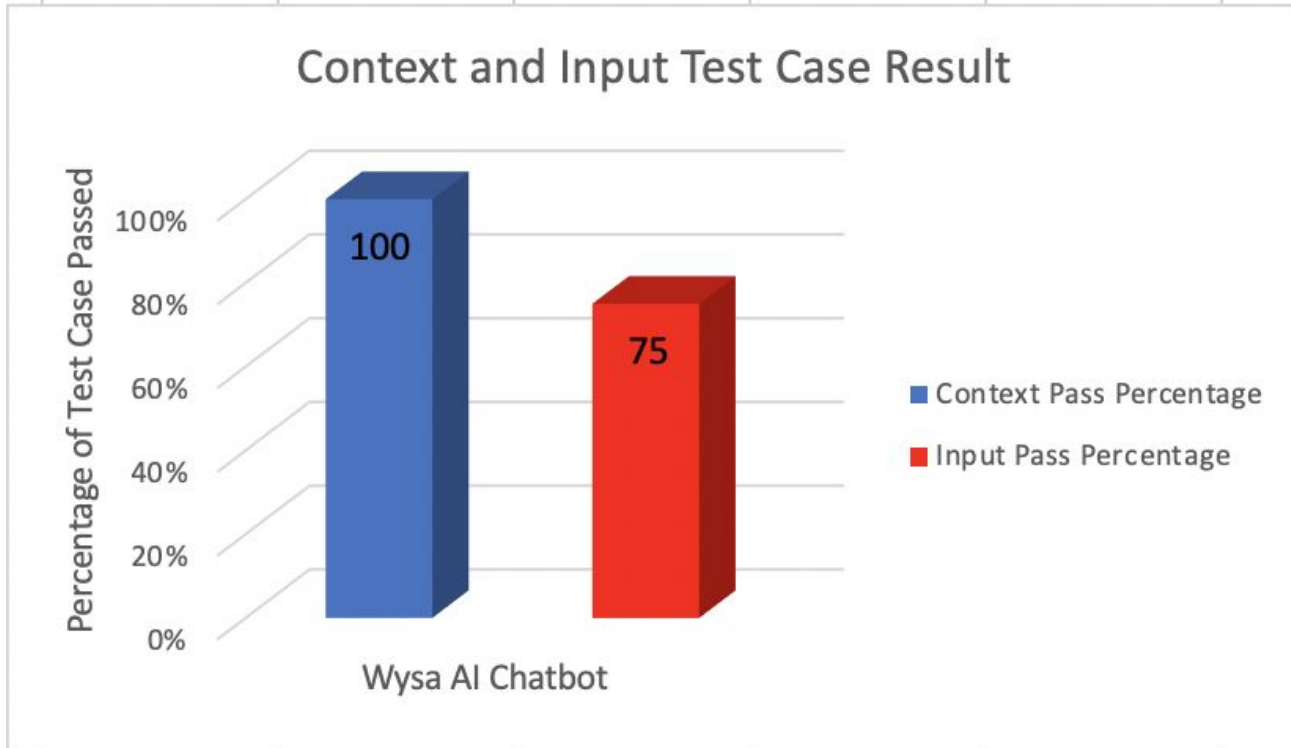
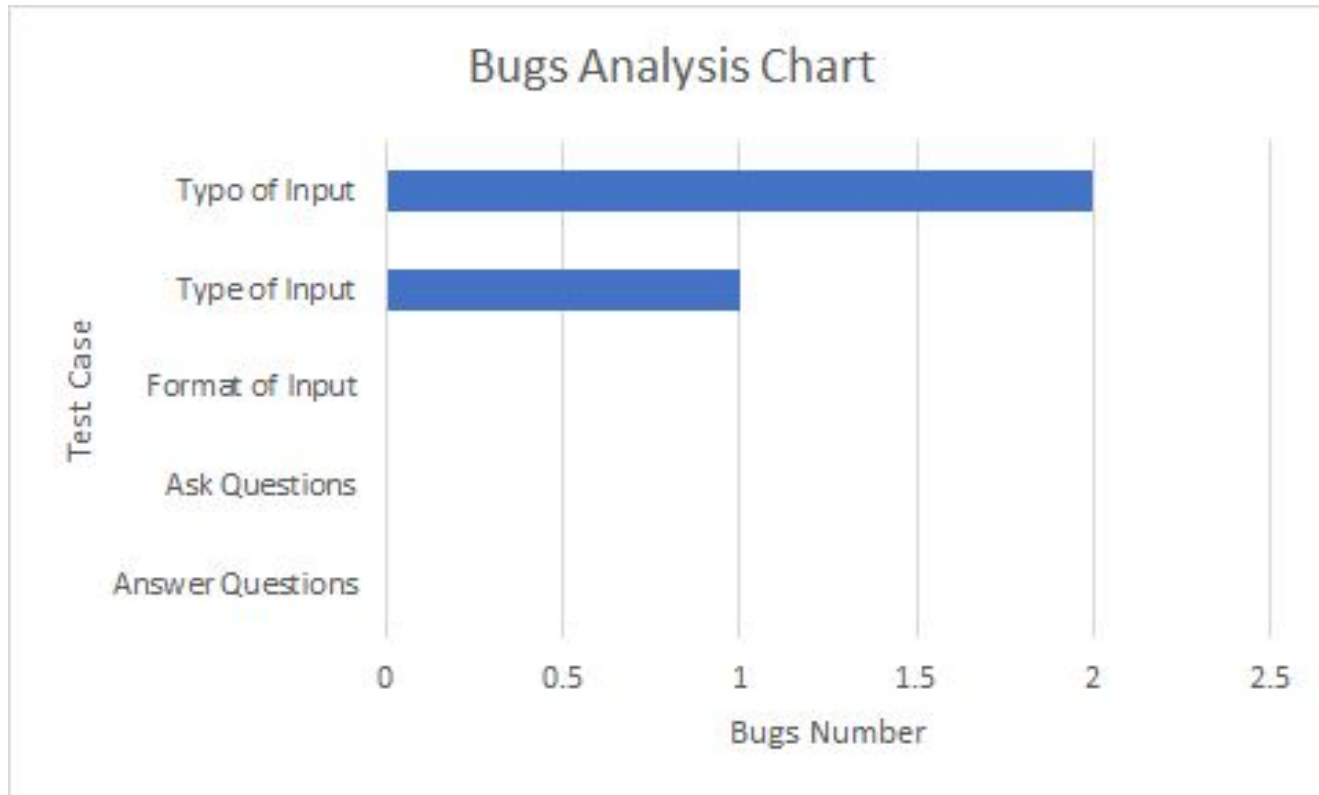# Test Case Results

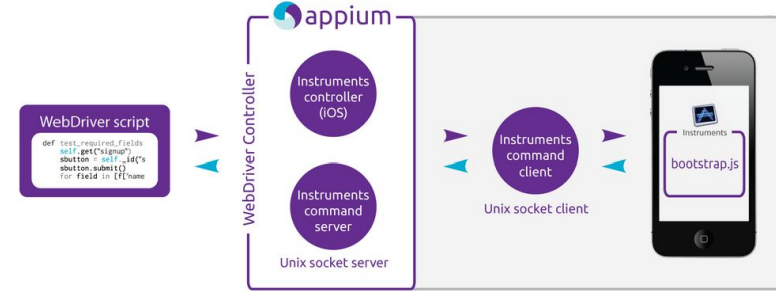# Overall Test Results

AI Bugs/Errors

- Inability to predict a typo input (with 2 letters messed up)
    - For example: if user have a conversation and type "Depresed" instead of "Depressed" , "Yees" instead of "yes" or "alrihgt" instead of "alright", Wysa will not accept or understand the conversation.

Bug Analysis Chart
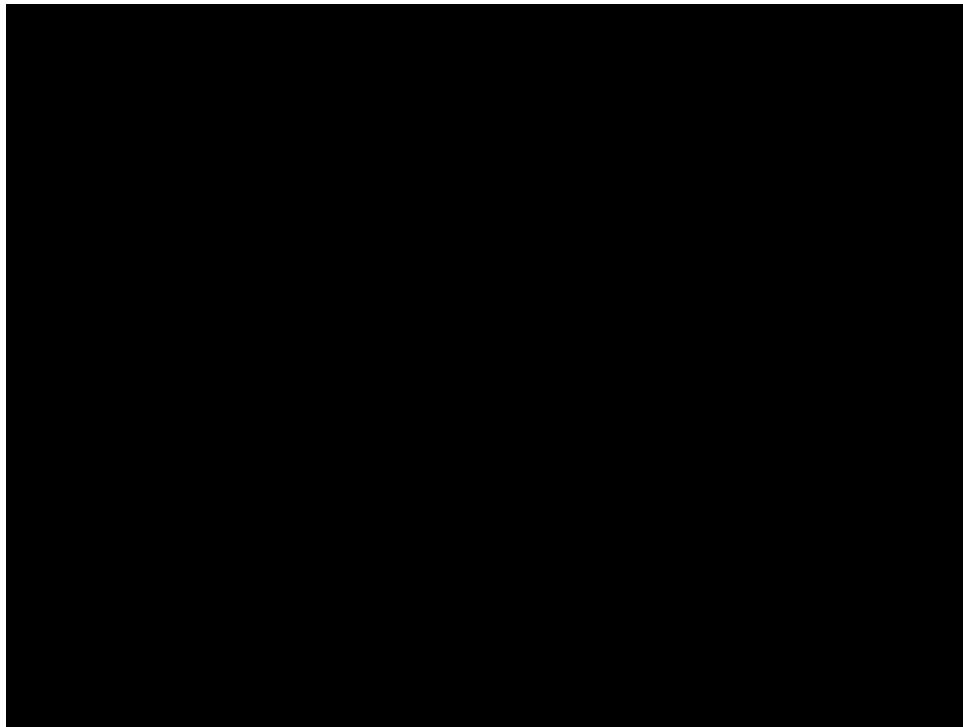
Test Automation and Strategy

For the Test Automation, our team uses Appium framework because Wysa is a mobile application that can run on Android platform.

The reason why we selected Appium because it has:

- Standard API: that is easier for us to test against Android platforms.
- Use any WebDriver compatible language: we can select any compatible language to create script and run test. We chose Java.

Project Demo

Test Automation and Strategy (cont.)

Based on our tool and findings, because this is a chatbot AI app, we are not sure if automation would help testing due to receiving different answers.

There is some uncertainty if we were to leave testing with automation.

Conclusion

Based on our results, Wysa has a solid AI and we were very impressed. However, there is always room for improvement for this app based on our findings such as:

- Add the ability to memorize user's input for a period of time (because many mental health issue lead to mood swings and emotion swings)
- Add the ability to predict user's input typo