



Wysa AI Testing Analysis

GROUP 2
Khai Nguyen
Faith Deanon
Kevin Campbell
Long Nguyen

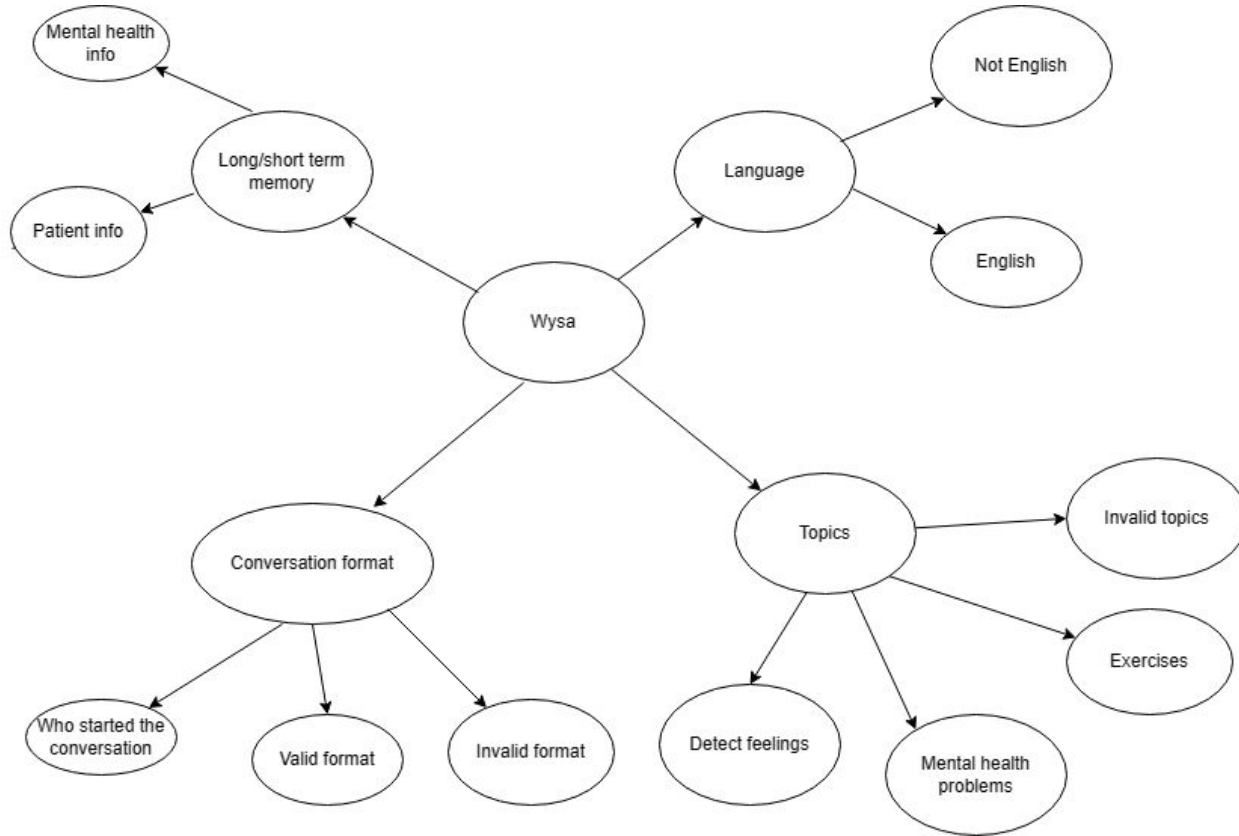


Introduction

We will be testing Wysa, an everyday mental health chat bot.



Test 4 main topics



- Detect feelings from user
- Detect mental health problems from user
- Walkthrough exercises
- In addition, we performed test cases to test other functionalities like memory, conversation format and language



Conventional Testing Method

- Decision table testing
- Equivalent testing
- Scenario testing

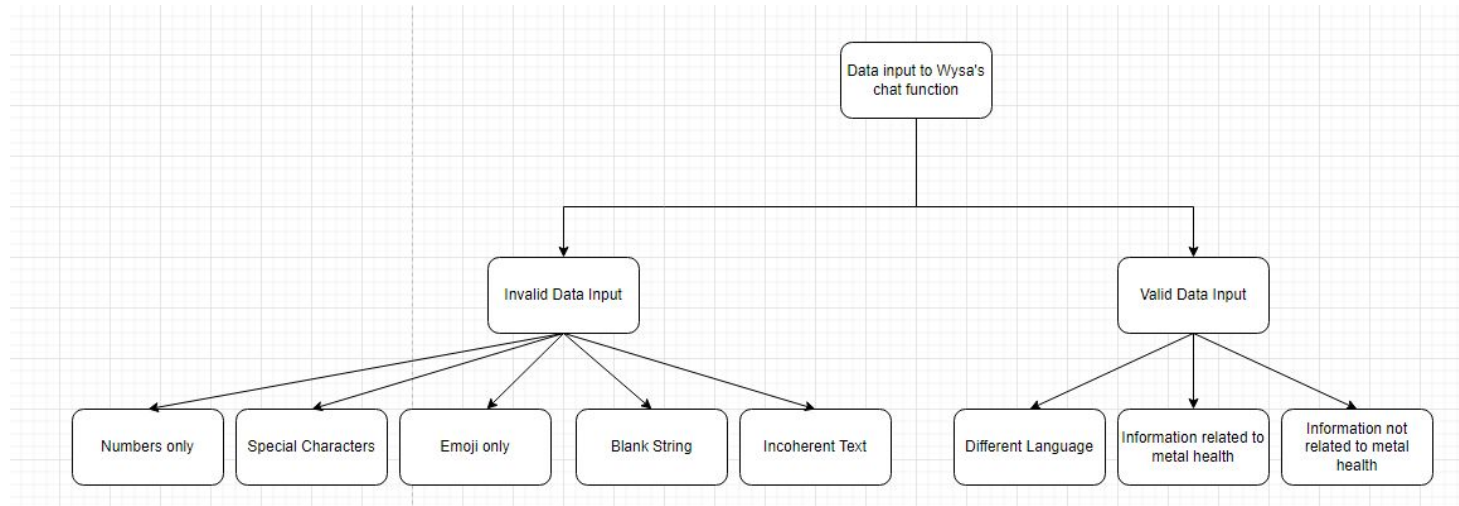


Decision table method

- Test with different combination of user's input and Wysa's conditions
 - Is your answer related to the current conversation?
 - Does the answer express any feelings?
 - Does the answer change what you said previously?
 - Is the answer one of the offered responses by Wysa?

Equivalent testing

- Test input in different format such as numbers, invalid grammar, etc





Scenario testing

- Test by acting as user through different scenario such as
 - User uses the app to call SOS
 - To express they have a good day
 - To express they have a bad day
 - To go through exercises



Conventional Testing Method results

Test Complexity

- Decision table: 16
- Equivalence partitioning test: 8
- Scenario test: 4

Result (Passed rate)

- Decision table: 75%
- Equivalence partitioning test: 87.5%
- Scenario test: 100%



Conventional Testing Method test costs

| Test | Cost |
|-----------------------|-------------|
| Decision table | 360 minutes |
| Equivalence partition | 300 minutes |
| Scenario testing | 240 minutes |



Conventional test method bug analysis

Bugs

- Wysa often assume what user is saying when it asks user to elaborate on things or middle of the conversation

Feeling detection

- Wysa is good at detecting feelings in sentences as well when users change their feelings and can offer help.
- Better when the user is answered using one of the offered responses, if not then it is better for the user to keep their sentences as concise as possible



Conventional test method bug analysis (2)

Conversation flow

- Wysa does not offer much conversation variety
- Often assumes what user was saying when the user is in the middle of conversation
- To change the subject effectively, it is better to end the current conversation first or signal Wysa you want to say something else instead of doing it abruptly

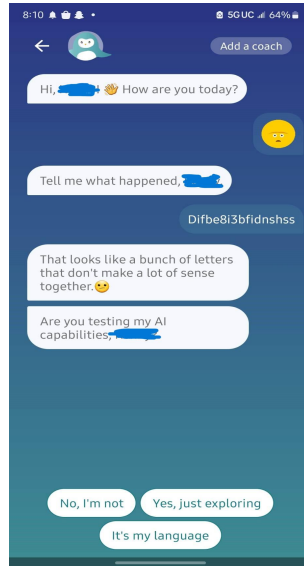
Invalid inputs

- Wysa often ask users if they are testing the AI's capabilities if it detects invalid inputs
- Wysa does not allow blank strings

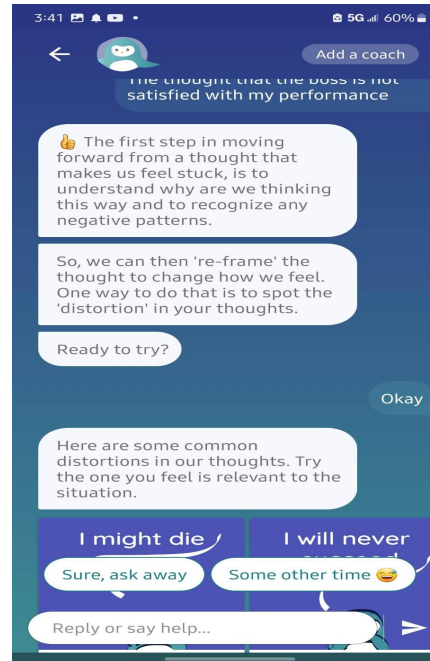
Valid Inputs

- Wysa doesn't understand any other language than English

Some pictures



- Testing whether or Wysa recognizes bunch of letters that don't make sense



- User going through the re-frame thought exercise in Wysa
- Reframe thoughts is the exercise that Wysa offers to help user view negative past events in a positive way



AI Testing method

We will be testing:

1. Mental illness function
2. Exercise function
3. Feelings detection function
4. Invalid topics



AI Testing Method Access

Mental Illness Function

Username: afdeanon

Password: Tromb0n3_pwd

AI WYSA Testing -> Mental Illness or Problems Models Del 2A 10



AI Testing Method Access

Exercise function:

Username: khainl01

Password: khainl01

Exercisefunction9

Invalid Topics function:

Username: Kevin_521

Password: 5223

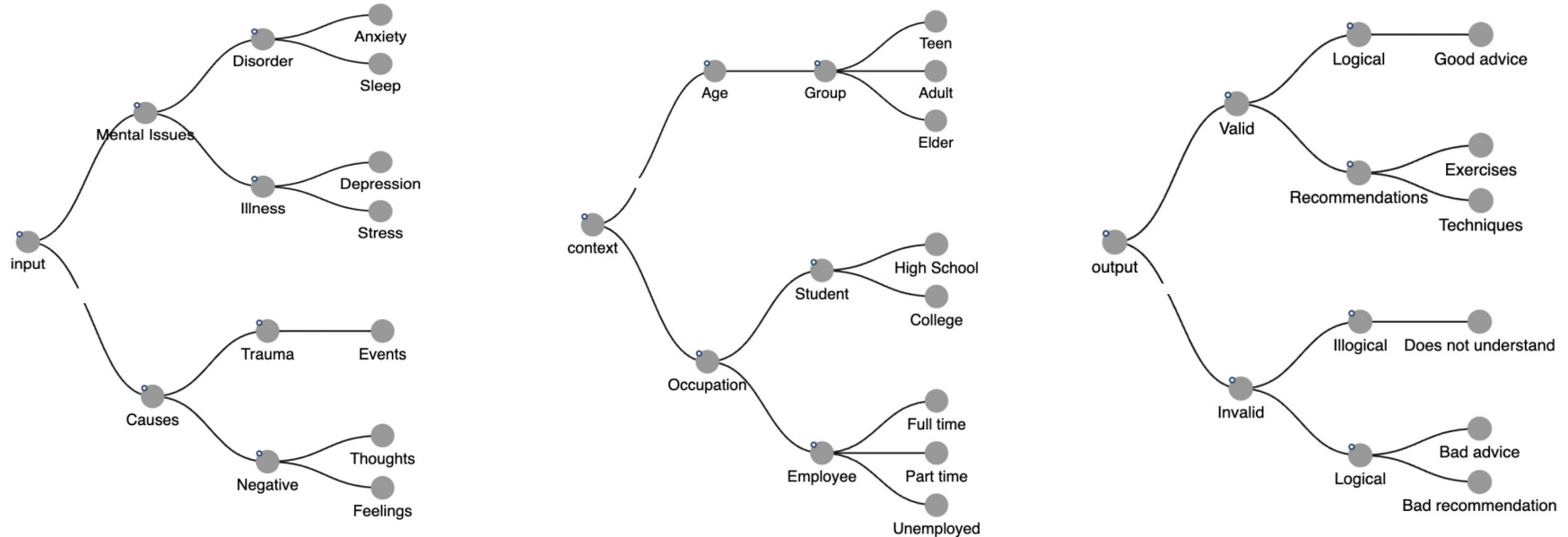
Feeling Detection function

Username: longpn0504

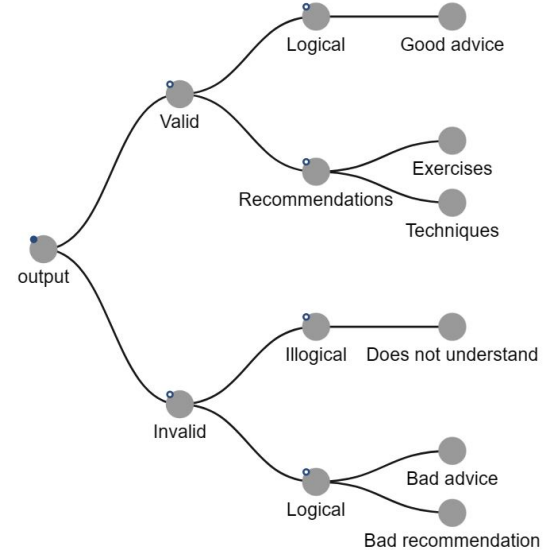
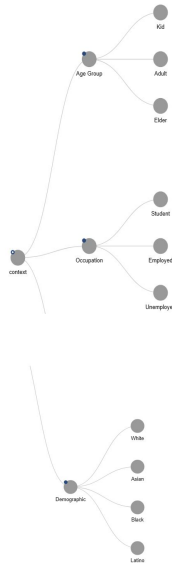
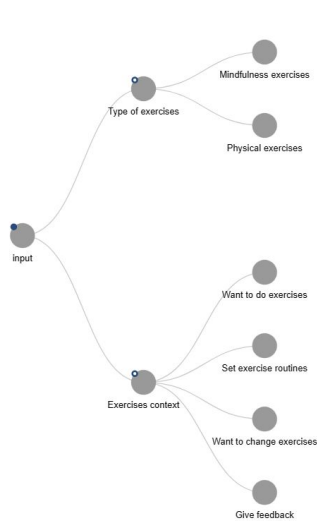
Password: longpn0504

AI Testing Project-Wysa

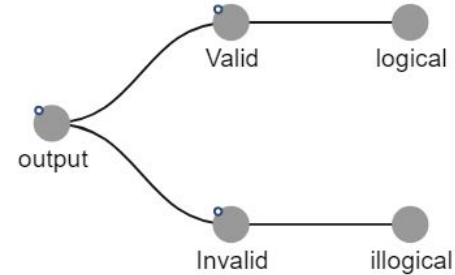
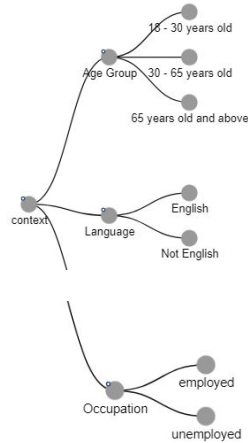
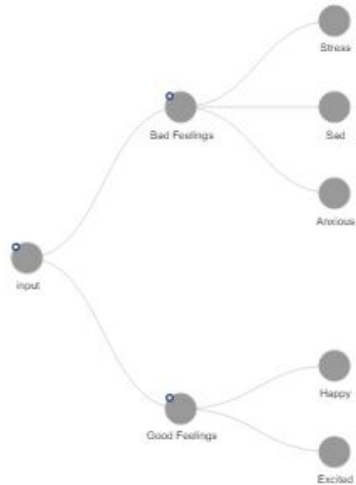
Mental Illness Input, Context Model, and Output



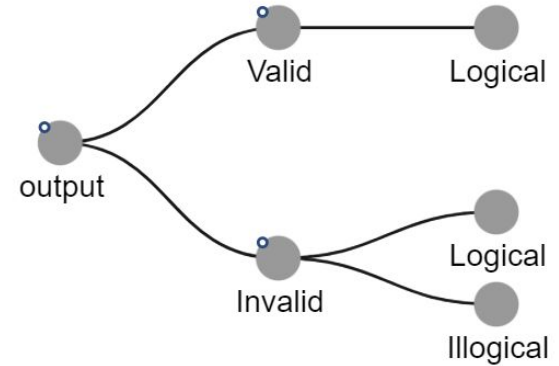
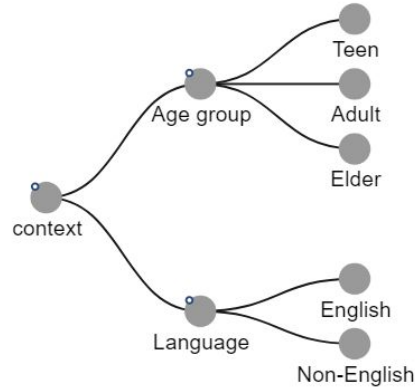
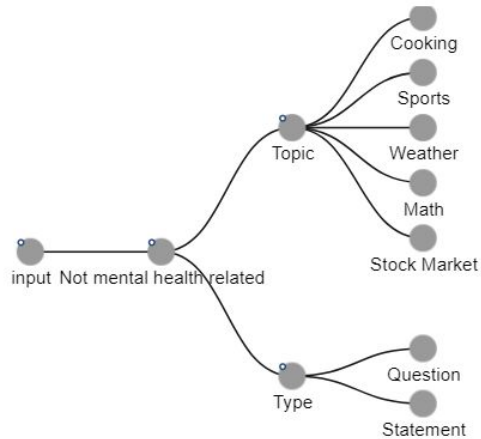
Exercises function input, context model and output



Feelings Detection Input, Context, and Output Model



Invalid Topics Input, Context, and Output Model



Mental Illness Decision Table

| Id | Age.Group | Occupation.Employee | Occupation.Student |
|----|-----------|---------------------|--------------------|
| 1 | Teen | Full time | High School |
| 2 | Teen | Part time | High School |
| 3 | Teen | Unemployed | High School |
| 4 | Teen | Full time | College |
| 5 | Teen | Part time | College |
| 6 | Teen | Unemployed | College |
| 7 | Adult | Full time | High School |
| 8 | Adult | Part time | High School |
| 9 | Adult | Unemployed | High School |
| 10 | Adult | Full time | College |

| Id | Invalid.Illlogical | Invalid.Logical | Valid.Logical | Valid.Recommendations |
|----|--------------------|-----------------|---------------|-----------------------|
| 1 | undefined | undefined | undefined | undefined |
| 2 | undefined | undefined | undefined | undefined |
| 3 | undefined | undefined | undefined | undefined |
| 4 | undefined | undefined | undefined | undefined |
| 5 | undefined | undefined | undefined | undefined |
| 6 | undefined | undefined | undefined | undefined |
| 7 | undefined | undefined | undefined | undefined |
| 8 | undefined | undefined | undefined | undefined |
| 9 | undefined | undefined | undefined | undefined |
| 10 | undefined | undefined | undefined | undefined |

| Id | Invalid.Illlogical | Invalid.Logical | Valid.Logical | Valid.Recommendations | Id | Causes.Negative | Causes.Trauma | Mental Issues.Disorder | Mental Issues.Illness |
|----|--------------------|-----------------|---------------|-----------------------|----|-----------------|---------------|------------------------|-----------------------|
| 1 | undefined | undefined | undefined | undefined | 1 | Thoughts | Events | Anxiety | Depression |
| 2 | undefined | undefined | undefined | undefined | 2 | Feelings | Events | Anxiety | Depression |
| 3 | undefined | undefined | undefined | undefined | 3 | Thoughts | Events | Anxiety | Stress |
| 4 | undefined | undefined | undefined | undefined | 4 | Feelings | Events | Anxiety | Stress |
| 5 | undefined | undefined | undefined | undefined | 5 | Thoughts | Events | Sleep | Depression |
| 6 | undefined | undefined | undefined | undefined | 6 | Feelings | Events | Sleep | Depression |
| 7 | undefined | undefined | undefined | undefined | 7 | Thoughts | Events | Sleep | Stress |
| 8 | undefined | undefined | undefined | undefined | 8 | Feelings | Events | Sleep | Stress |

Exercise Decision Table

| Id | Conversation.Do exercise | Conversation.Feedback | Exercises.Mental | | Exercises.Physical |
|----|--------------------------|-----------------------|------------------|-------------|--------------------|
| | | | Reframe thoughts | Mindfulness | Strength |
| 1 | One time | Good | | | |
| 2 | One time | Bad | | | |
| 3 | Set exercise schedule | Good | | | |
| 4 | Set exercise schedule | Bad | | | |
| 5 | One time | Good | | | |
| 6 | One time | Bad | | | |
| 7 | Set exercise schedule | Good | | | |
| 8 | Set exercise schedule | Bad | | | |
| 9 | One time | Good | | | |
| 10 | Set exercise schedule | Good | | | |

| Id | Invalid.Illlogical | Invalid.Logical | Valid.Logical | Valid.Recommendations | Exercises |
|----|---------------------|-----------------|---------------|-----------------------|-----------|
| | | | | | |
| 1 | Does not understand | Bad advice | Good advice | | |
| 2 | undefined | undefined | undefined | undefined | |
| 3 | undefined | undefined | undefined | undefined | |
| 4 | undefined | undefined | undefined | undefined | |
| 5 | undefined | undefined | undefined | undefined | |
| 6 | undefined | undefined | undefined | undefined | |
| 7 | undefined | undefined | undefined | undefined | |
| 8 | undefined | undefined | undefined | undefined | |
| 9 | undefined | undefined | undefined | undefined | |
| 10 | undefined | undefined | undefined | undefined | |

| Id | Age Group | Demographic | Occupation |
|----|-----------|-------------|------------|
| | | | |
| 1 | Kid | White | Student |
| 2 | Kid | Asian | Student |
| 3 | Kid | Black | Student |
| 4 | Kid | Latino | Student |
| 5 | Kid | White | Student |
| 6 | Kid | Asian | Employed |
| 7 | Kid | Black | Employed |
| 8 | Kid | Latino | Employed |
| 9 | Kid | White | Employed |
| 10 | Kid | Asian | Unemployed |

Feelings Topic Decision Table

| Demographic Attributes | | | Sentiment Attributes | | | Specific Feelings | | |
|------------------------|-------------|------------|----------------------|-------------|--------|-------------------|---------|---------|
| Age Group | Language | Occupation | Good Feeling | Bad Feeling | Strong | Happy | Excited | Relaxed |
| 18-24 years old | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 25-34 years old | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 35-44 years old | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 45-54 years old | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 55-64 years old | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 65 years and above | English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 18-24 years old | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 25-34 years old | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 35-44 years old | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 45-54 years old | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 55-64 years old | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |
| 65 years and above | Not English | employed | employed | unemployed | Strong | Happy | Excited | Relaxed |

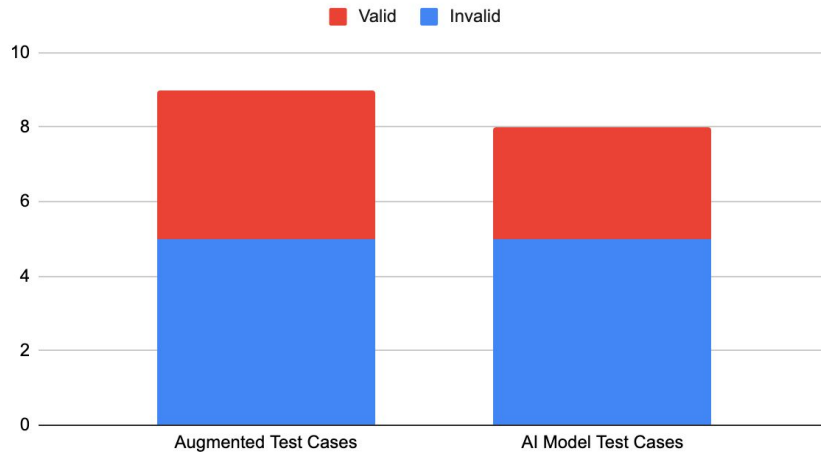
Invalid Topic Decision Table

| id | Invalid | | Valid | | id | Not mental health related Topic | | Not mental health related Type |
|----|-----------|-----------|-----------|-----------|----|---------------------------------|-----------|--------------------------------|
| | undefined | undefined | undefined | undefined | | undefined | undefined | |
| 1 | undefined | undefined | undefined | undefined | 1 | Cooking | Question | |
| 2 | undefined | undefined | undefined | undefined | 2 | Stock Market | Statement | |
| 3 | undefined | undefined | undefined | undefined | 3 | Cooking | Question | |
| 4 | undefined | undefined | undefined | undefined | 4 | Sports | Statement | |
| 5 | undefined | undefined | undefined | undefined | 5 | Sports | Question | |
| 6 | undefined | undefined | undefined | undefined | 6 | Weather | Statement | |
| 7 | undefined | undefined | undefined | undefined | 7 | Weather | Question | |
| 8 | undefined | undefined | undefined | undefined | 8 | Math | Statement | |
| 9 | undefined | undefined | undefined | undefined | 9 | Math | Question | |
| 10 | undefined | undefined | undefined | undefined | 10 | Stock Market | Question | |

| id | Age group | Language |
|----|-----------|-------------|
| 1 | Teen | English |
| 2 | Teen | Non-English |
| 3 | Adult | English |
| 4 | Adult | Non-English |
| 5 | Elder | English |
| 6 | Elder | Non-English |
| | | |

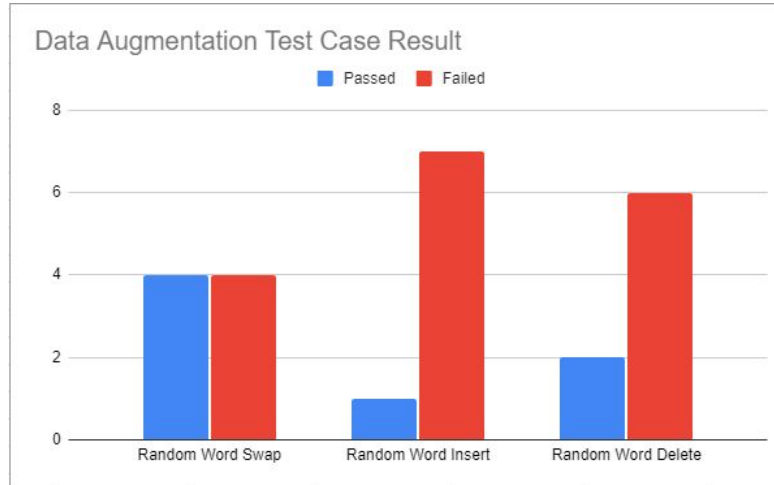
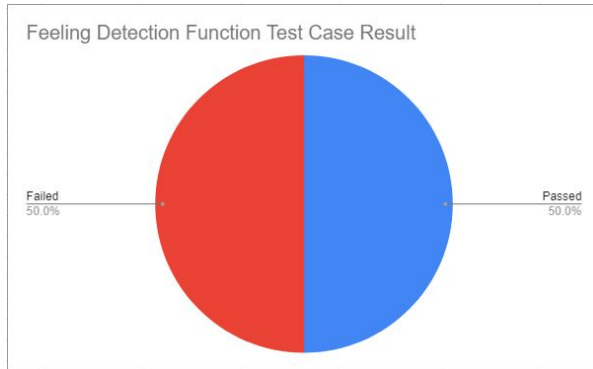
Mental Illness Test case analysis

Mental Illness Test Case Analysis



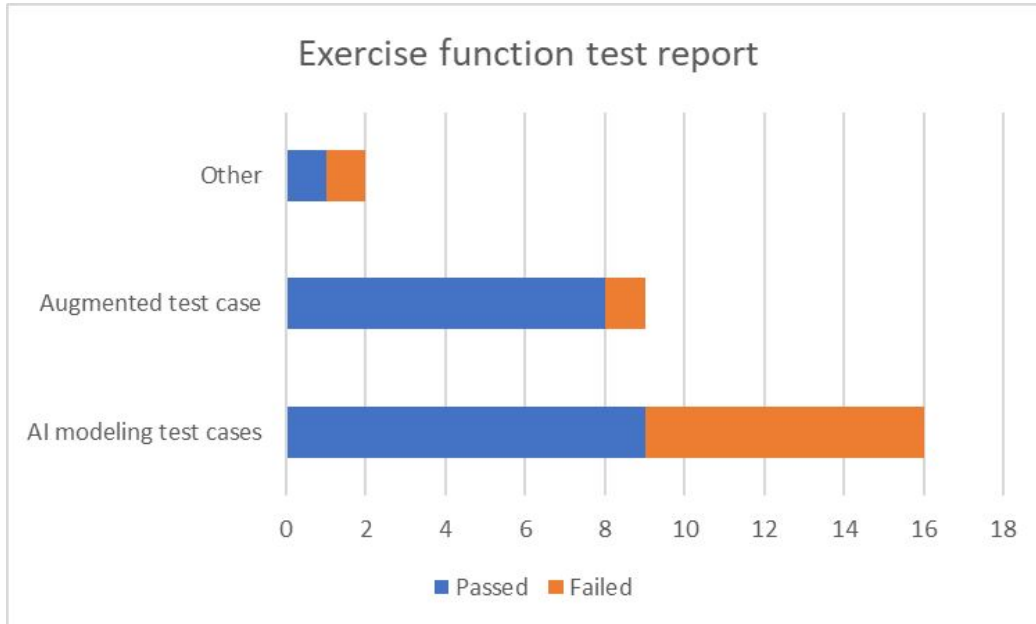
- There are 8 test cases, 9 augmented test cases
- The test complexity is 17
- Wysa fails 55.55% of the text data augmentation test cases
- Wysa fails 62.5% of the AI Model test cases

Test case analysis (Feeling function)



- There are 32 test cases
- Wysa failed 4 out of 8 AI, 50% powered test cases
- Wysa failed 7 out of 24, 29.16% augmented test cases

Test case analysis (Exercises function)

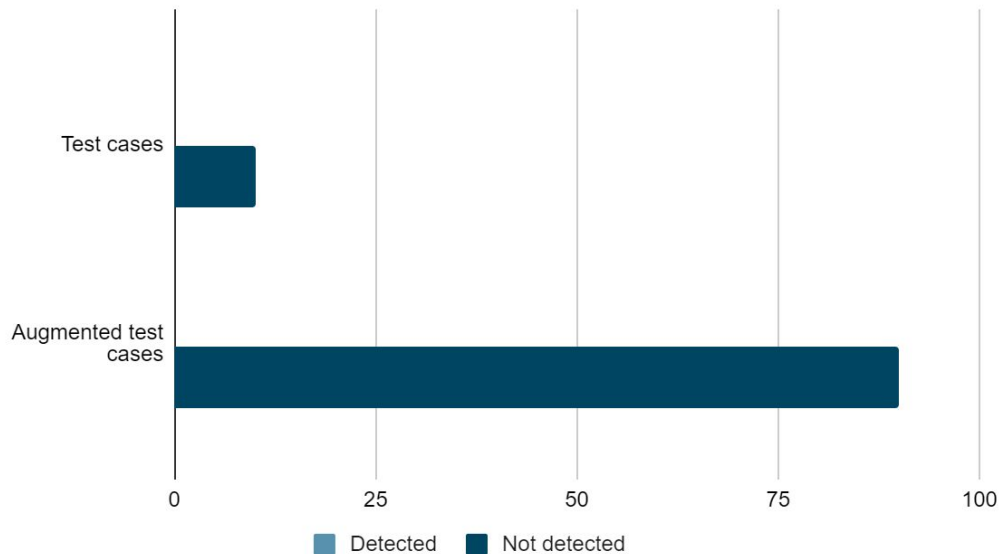


- There are 9 augmented test cases, 16 AI modeling test cases and 2 test cases test conversation format and different language
- Other test cases passed 1 out of 2, 50%
- Augmented test case passed 8 out of 9, 88.89%
- AI modeling test cases passed 9 out of 16, 56.25%



Test case analysis (Invalid function)

Invalid Topic Detection



Test complexity: 100

Base test cases: 10

Augmented: 90

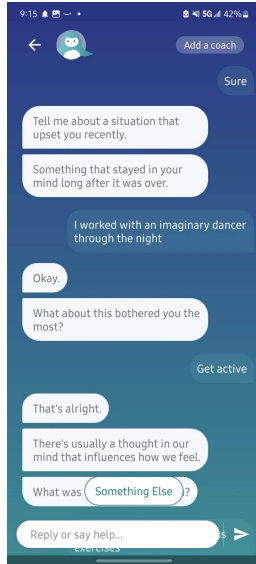
Wysa failed all the test cases



Bug reporting

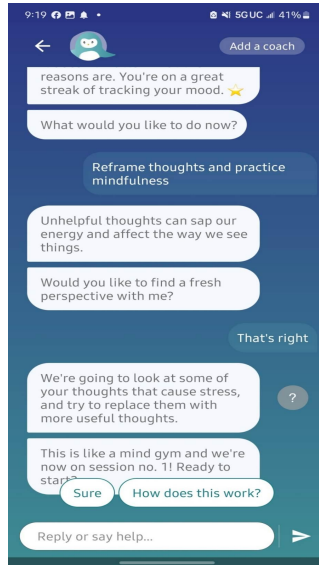
- Wysa is good at detect user feelings and if user input conflicted feelings, Wysa can detect it
- Wysa exercises function is very basic and only covers walkthrough mindfulness exercises
- Wysa is not able to understand different languages and can only respond in english
- When User doesn't follow Wysa's conversation format, Wysa will provide irrelevant suggestions/recommendations or will not understand the User's needs
- Wysa was able to remember short term mental health and causes inputs and long term mental health and causes inputs

Some bug examples



- Wysa ignored user request to 'Get active' and continued the conversation

Some bug examples



- User requested 2 different things, 'reframe thoughts' and 'practice mindfulness'
- But Wysa assumed that user wanted to reframe thoughts only

Some bug examples

- User requested 2 different things, 'reframe thoughts' and 'get active'
- But Wysa assumed that user wanted to reframe thoughts only

