

WYSA AI TEST SCRIPTS

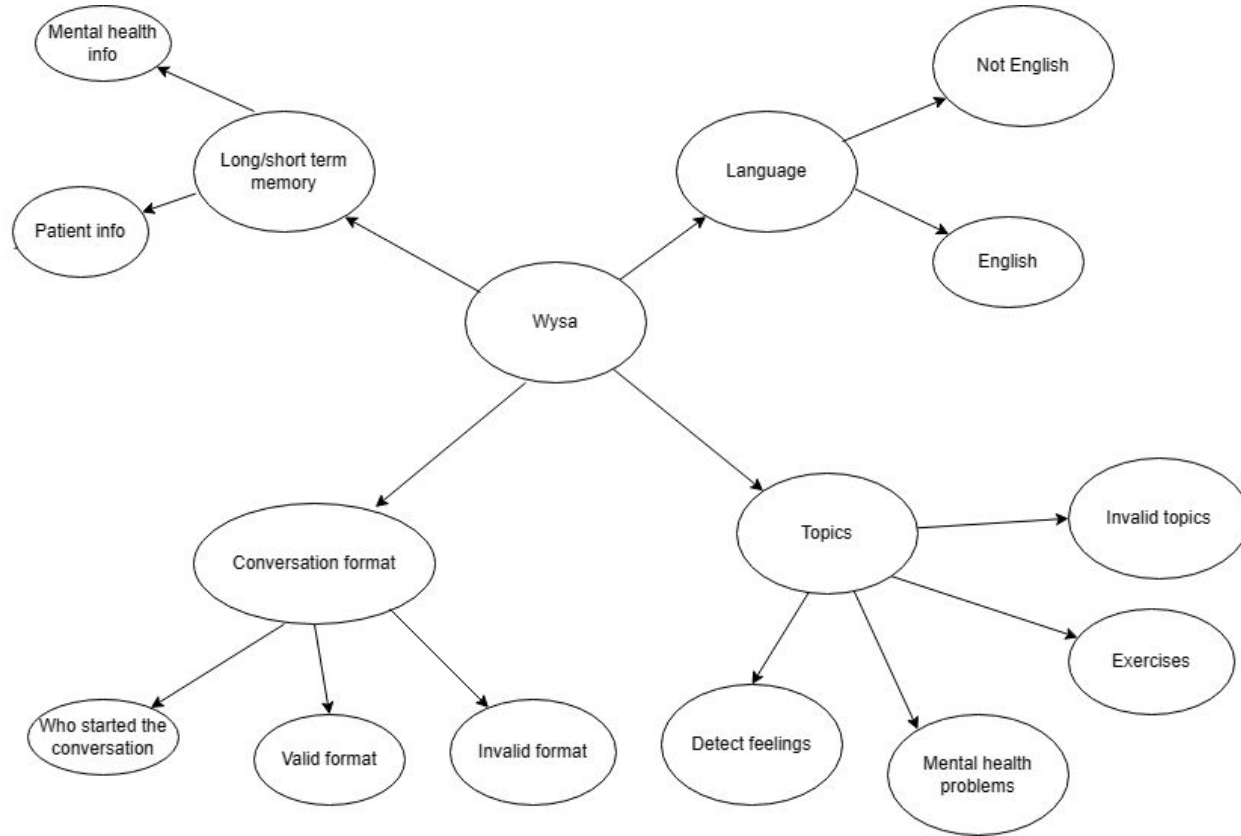
GROUP 2
Khai Nguyen
Faith Deanon
Kevin Campbell
Long Nguyen

INTRODUCTION

- We developed test scripts for Wysa app
- Platform: Android



TEST 4 MAIN TOPICS



- We test 4 main topics: Invalid topics, Exercises, Mental health problems and Detect feelings
- We also test features such as memory, conversation format and different language

TOOLS AND TECHNOLOGIES USED

- Appium as testing server
- Android Emulator
- Javascript as scripting language

SCRIPTING STEPS FOR AUTOMATE TEST CASES

1. Test script reads the inputs and expected outputs
2. Test script enters the test input to Wysa
3. Test script collects the output
4. Test script performs similarity score
5. Test script writes result to a file
6. Repeat until done

TEST SCRIPT

```
// tap on the skip button
// let el21 = await driver.$("/hierarchy/android.widget.FrameLayout/android.widget.LinearLayout/android
// await el21.click());

await driver.pause(5000);
let textInput = "Let's get active " + Math.floor(Math.random() * 50);
let expectedTextOutput = "Let's try some exercises";

console.log(textInput);
let el2 = await driver.$("//android.widget.LinearLayout[@content-desc='Continue conversation']/andr
await el2.click();
await driver.pause(5000);
let el3 = await driver.$("~Reply or say help...");
await el3.setValue(textInput);
let el4 = await driver.$("~Send");
await el4.click();

await driver.pause(10000);

const elements1 = (await driver.$$('android.widget.TextView')).slice(-7);
let bagOfTexts = null;
```

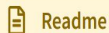
EXCEL INPUT FILE

Input	Expected output
Give some me mindfulness exercises for Asian elders	Would you like to try some exercises right now?
me some exercises for Asian elders	Would you like to try some exercises right now?
Give heedfulness me some mindfulness exercises for Asian elders	Would you like to try some exercises right now?
Afford me some heedfulness exercises for Asian elders	Would you like to try some exercises right now?
I want to reframe thoughts to feel better as a L	Would you like to try some exercises right now?
Give me some mindfolne88 exercises for Asian	Would you like to try some exercises right now?
Give me some mindfulness exe3dises for Asian #lders	Would you like to try some exercises right now?
wGive me some mindfulness exercises for A2s	Would you like to try some exercises right now?
Gvie me some mindfulness exercises for Asian eldres	Would you like to try some exercises right now?
Give me som mindfulness exercises for Asian elers	Would you like to try some exercises right now?

COMPARISON

string-similarity-js 

2.1.4 • Public • Published 2 years ago



Readme



Code



Beta



0 Dependencies



Build Status



codecov

100%



wallaby.js

configured



code quality

A

String Similarity

Based on the **Sørensen-Dice coefficient**, this algorithm is most effective at detecting rearranged words or misspellings. It tends to be less effective with very short strings, unless perhaps you switch to comparing individual characters in common instead of bigrams.

It is case insensitive unless you specify otherwise. Does not ignore punctuation or spaces. In some cases, removing punctuation beforehand may improve accuracy.

- For string comparison, we used the library as specified on the right
- The library implemented Sorensen-Dice coefficient for similarity string checking

MULTIPLE TEST SCRIPTS RESULT FILE

Input: The reframe thought exercise was not helpful

Expected output: I am sorry I will strive to do better

Actual output: I sense you are finding this conversation unhelpful. Is that right, Hhmm?YesNot really, go on 🐼

Similarity score 0.16923076923076924

Passed? false

Input: Give me some strength exercises for Asian elders

Expected output: Would you like to try some exercises right now?

Actual output: Light physical activity is a great way to feel energised. 🏃 Would you like to try some quick exercises with me right now?

Let's go! 🐼 No, you misunderstood

Similarity score 0.4623115577889447

Passed? true

Input: I want to set some strength exercise routines as an Asian elder

Expected output: Would you like to try some exercise routines right now?

Actual output: You can explore the slider below for quick access to some of the features.You can also check your self-care kit for something that might help you.

Similarity score 0.3015075376884422

Passed? true

Input: This strength exercise was helpful for me

Expected output: Thanks for your feedback

Actual output: Light physical activity is a great way to feel energised. 🏃 Would you like to try some quick exercises with me right now?

Let's go! 🐼 No, you misunderstood

Similarity score 0.125

Passed? false

PROBLEMS

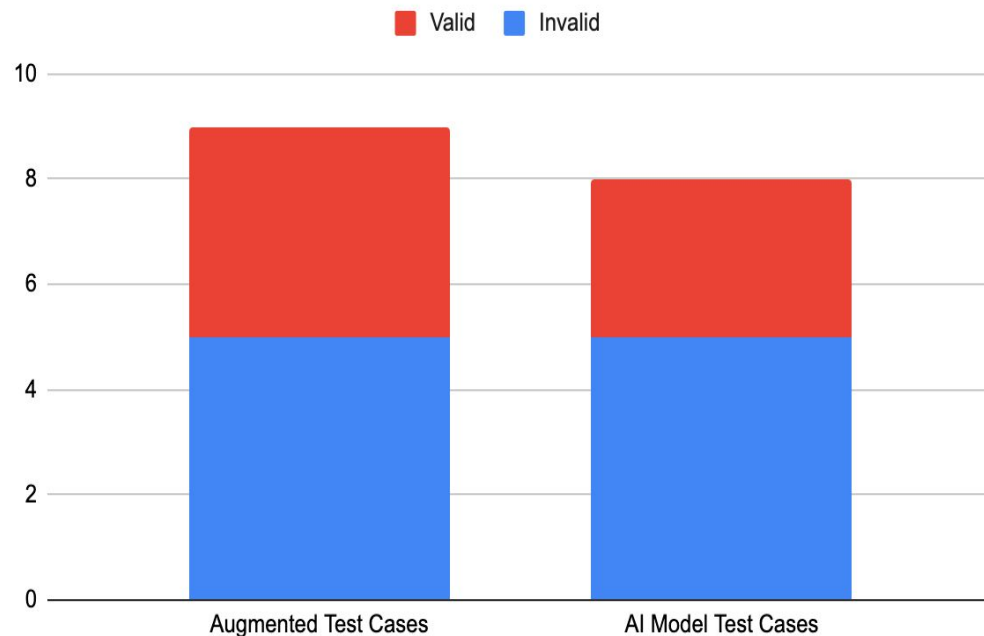
- Javascript doesn't have machine learning libraries like Python
- However, we had problems using Python because Javascript offers better asynchronous methods and code writing

TEST RESULT

Mental Illness function

- For this function, we automated a total of 17 test cases
 - We tested 9 data augmented test cases
 - 5 out of 9 test cases failed,
 - 44.44% success rate, 55.56% failure rate
 - We tested 8 AI powered test cases
 - 5 out of 8 test cases failed
 - 37.5% success rate, 62.5% failure rate

Mental Illness Test Case Analysis



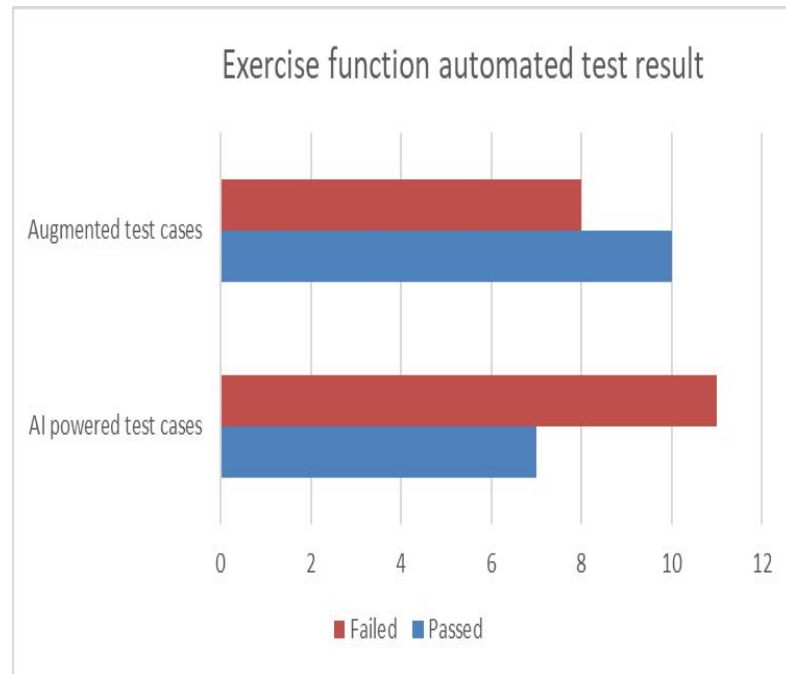
TEST RESULT

Exercise function

- For this functionality, we automated 36 test cases
 - 18 of them are AI powered test cases
 - At least one test case cover for each minor case in the input and context trees
 - 18 augmented test cases
 - Random word swap
 - Random word delete
 - Random word insert
 - Synonym augmentation
 - Ocr augmentation
 - Keyboard augmentation
 - Random char insert
 - Random char swap
 - Random char delete

Result is as below

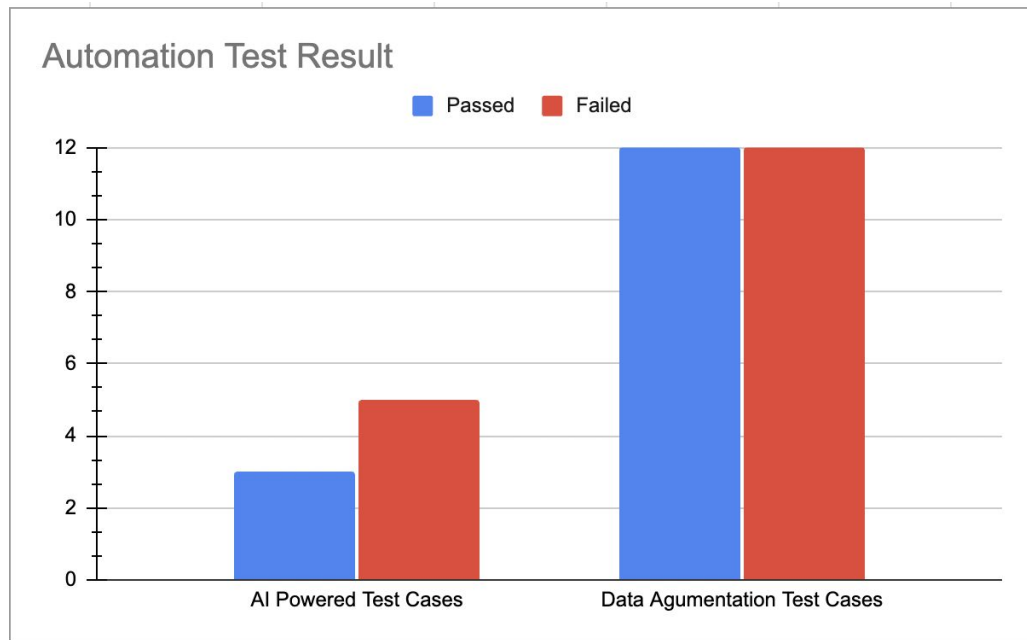
- AI powered test cases:
 - 7/18 passed, 38.89%
- Augmented test cases:
 - 10/18 passed, 55.56%



TEST RESULT

Feeling Detection Function

- We automated 32 test cases for this function
 - 8 of which are AI powered test cases
 - At least one test case cover for each minor case in the input and context spanning trees
 - 24 of which are data augmentation test cases
 - Random word insert
 - Random word delete
 - Random word swap
- Automation Test Result
 - AI powered test cases
 - 3/8 passed = 37.5%
 - Data augmentation test cases
 - 12/24 passed = 50%

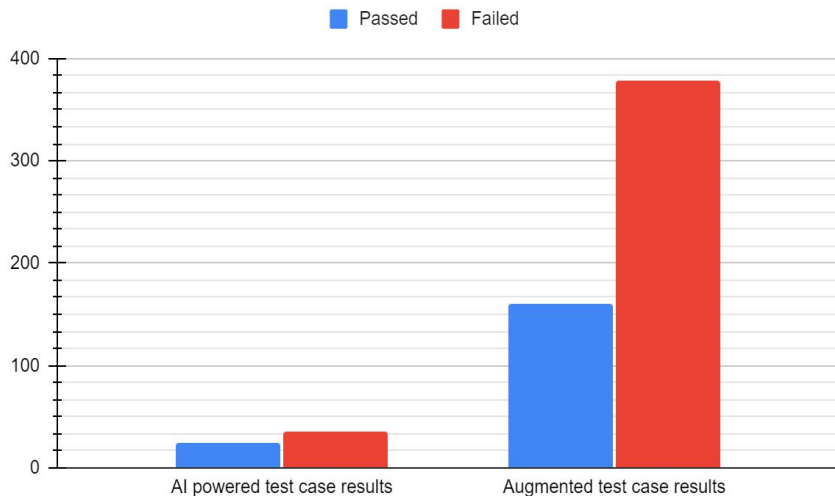


TEST RESULT

Invalid Topics

- This function automated 600 test cases
 - 60 of these are AI powered test cases
 - At least one test case for every input and context tree combo
 - 540 are augmented test cases
 - Random word swap
 - Random word delete
 - Random word insert
 - Synonym augmentation
 - Ocr augmentation
 - Keyboard augmentation
 - Random char insert
 - Random char swap
 - Random char delete
- Automated results
 - AI powered test cases
 - 24/60 passed = 40%
 - Augmented test cases
 - 161/540 passed = 29.8%

Invalid Topics Automated Test Result



TEST COST

Mental Illness function

- Each test takes about 20 seconds to complete
- The total time spent to test this function is $20 \times 17 = 360$ seconds, or 6 minutes

Exercise function

- Each test takes about 25 to 30 seconds to complete
- Therefore with 36 test cases, we spent a total of $30 \times 36 = 1080$ seconds, or 18 minutes

Feeling Detection Function

- Each test case take about 20 to 30 seconds to complete
- We spent about 14 minutes for this function

Invalid Topics

- Each test takes roughly 10 to 25 seconds to complete
- In total an average of 175 minutes, or almost 3 hours, is spent to test this function\

TEST COMPLEXITY

Test function	Test complexity
Mental Illness function	17
Exercise function	24
Feeling detection function	32
Invalid topics	600

Thank you