Deliverable 3 - AI Test Automation Report

CMPE 187 - Software Quality Engineering

San Jose State University

Group 2

Team members:

Khai Nguyen

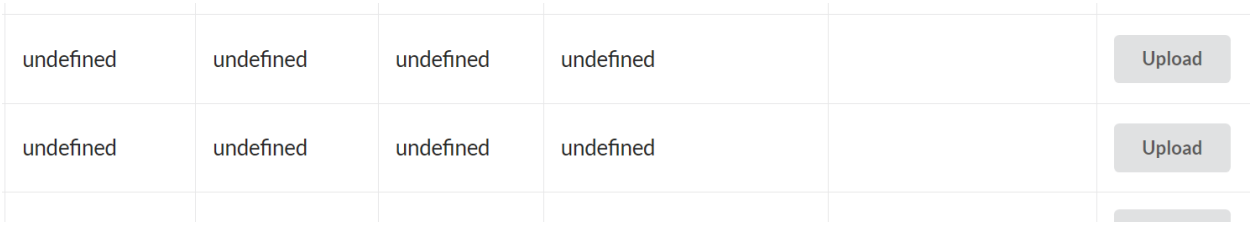
Faith Deanon

Kevin Campbell

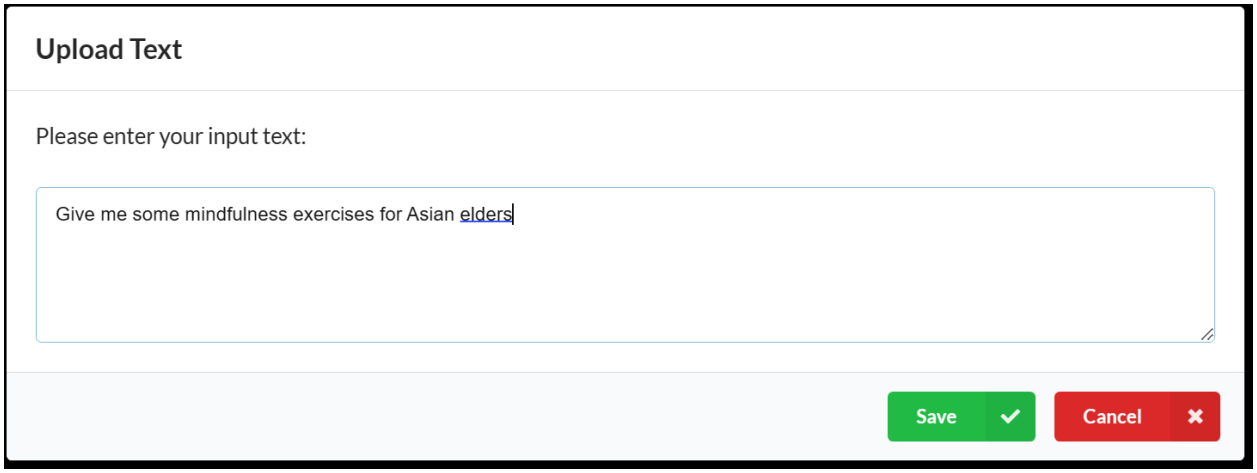
Long Nguyen

1. AI Testing Tool demo

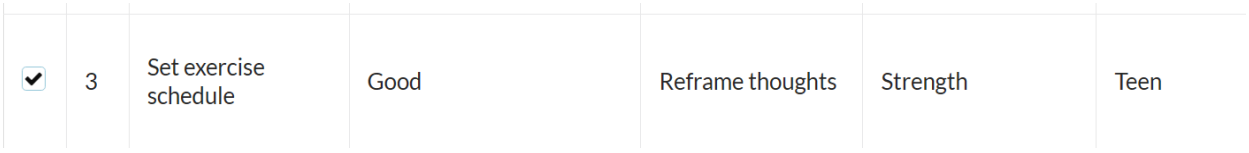
Data augmentation



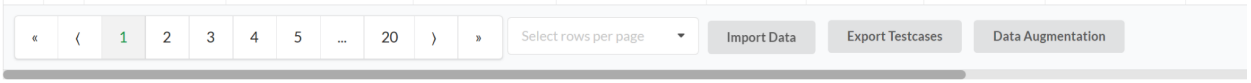
First click on Upload button for each test case



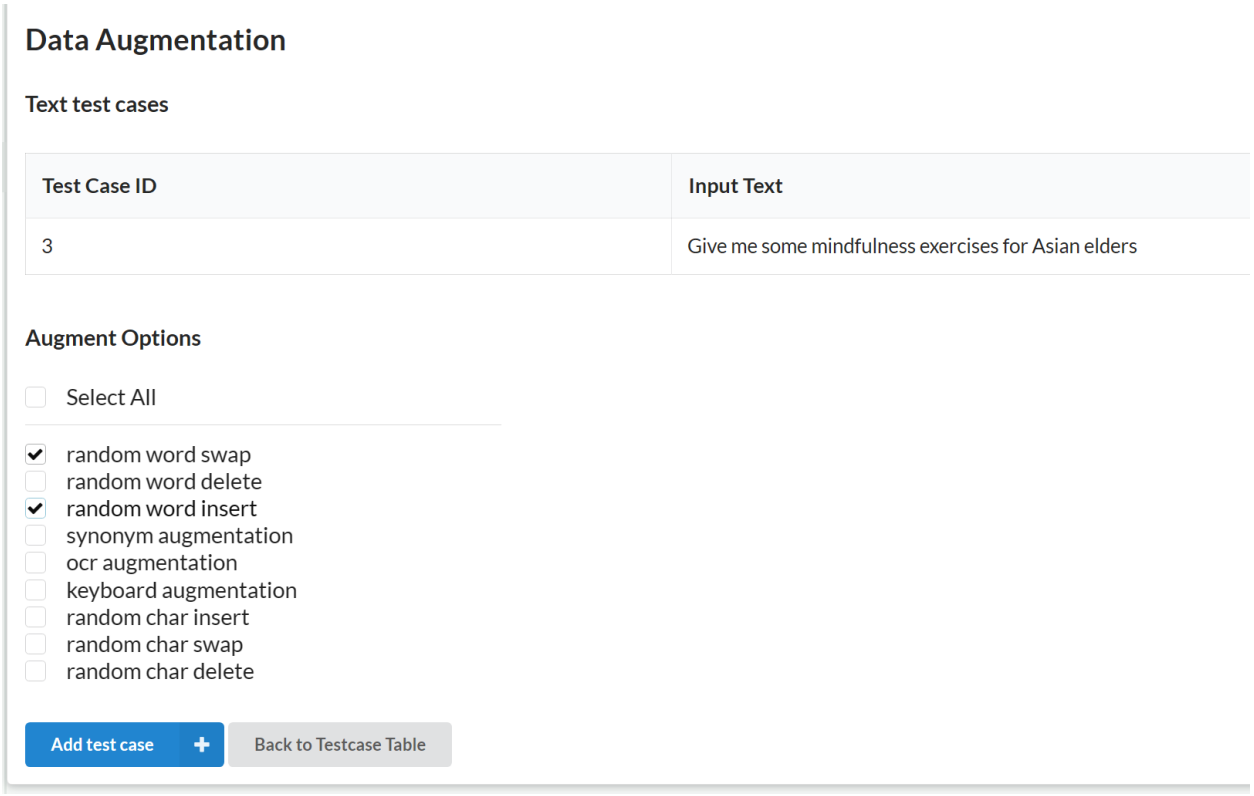
Then enter your text input



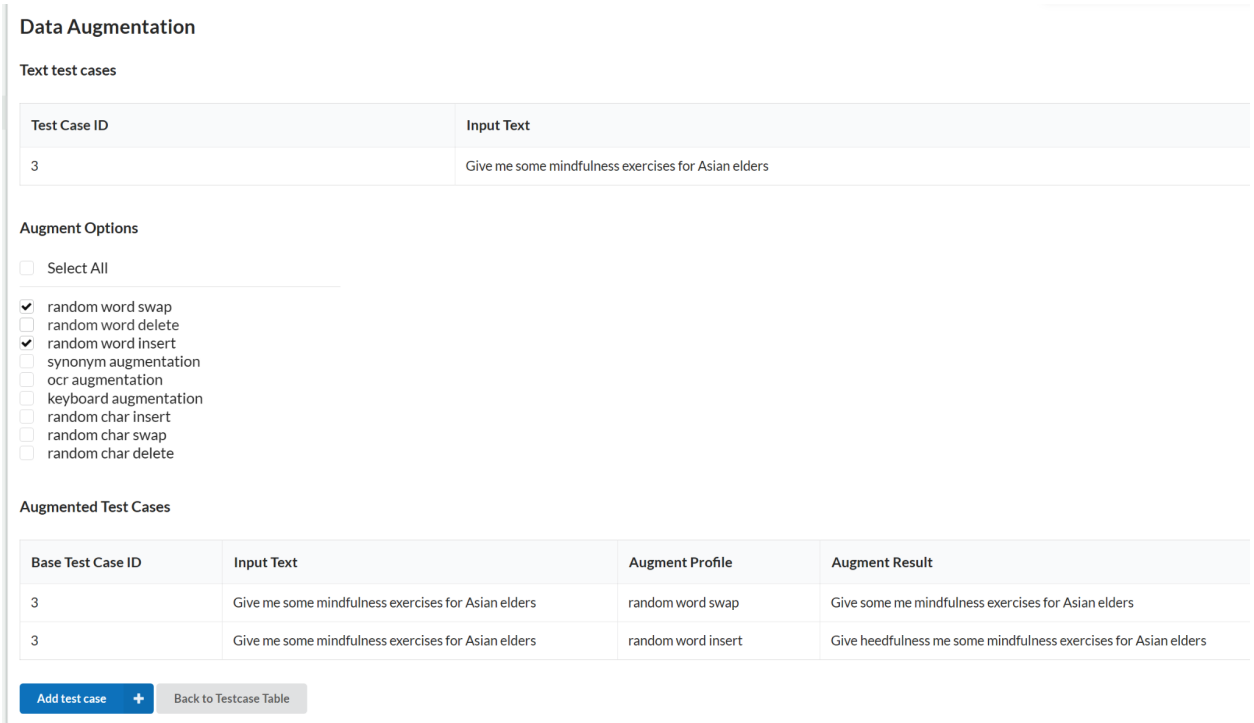
Then check the box on the test case



At below click on Data augmentation



Choose any data augmentation options



The tool will generate the augmented test case
Click on Back to Testcase Table to go back

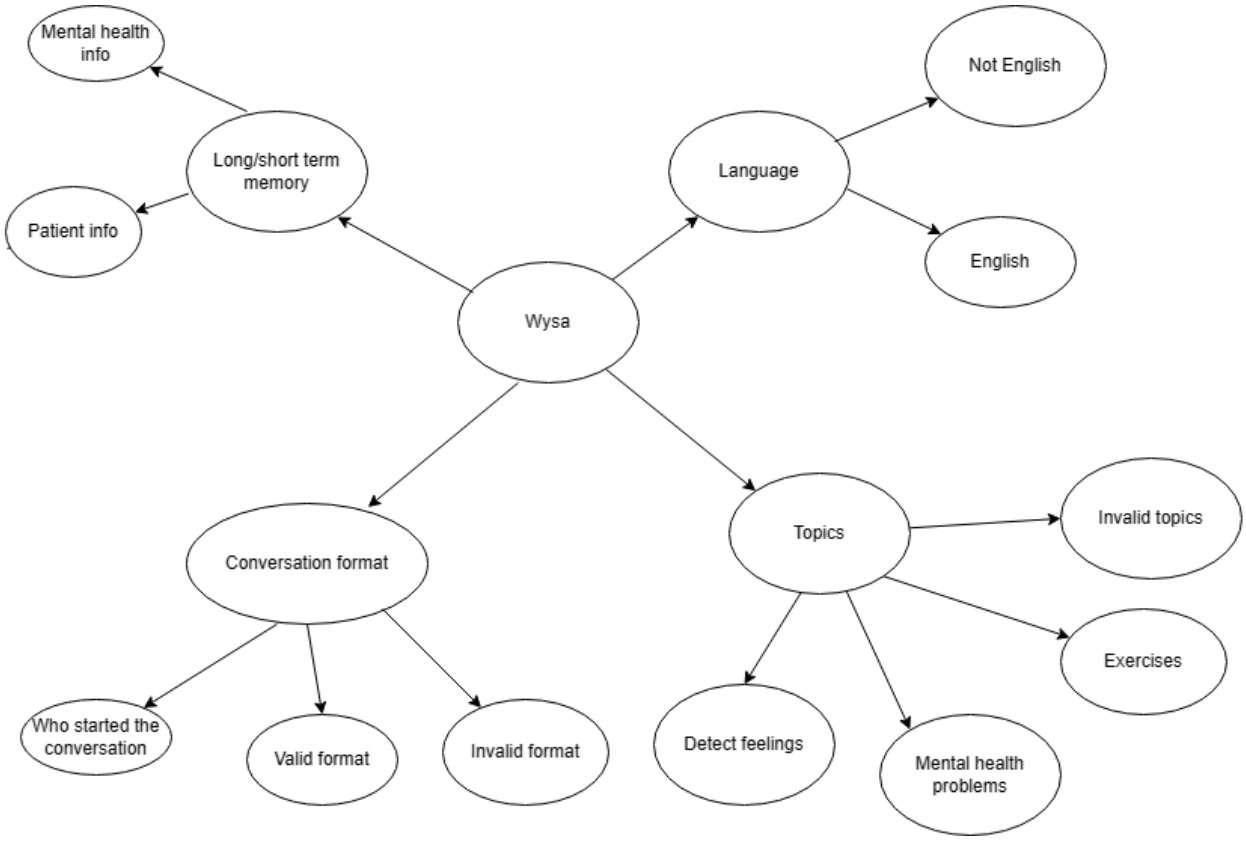
✔	3	Set exercise schedule	Good	Reframe thoughts	Strength	Teen

id: 1	augment method: random word swap
id: 2	augment method: random word insert

Everytime you check the text case, the augmented test case will be shown

2. Test scripts

We tested the Wysa app based on 4 topics:



- Those are Invalid topics, exercises topic, mental health problems and detect feelings topic
- We developed test script in order to test Wysa app, which is a chatbot that offer mental health counseling
- Our final version of test script supports
    1. Multiple test cases within one run
    2. The test script reads multiple input and expected output from Excel file
    3. Has similarity string comparison using Sorensen-Dice coefficient

Technologies:

- The test script used Javascript because the language offer good asynchronous method
- Runs on Android simulator
- We used Appium as testing server

3. Test scripts structure

1. In order to perform multiple test cases with just one run, our test scripts read all the input and expected output from Excel file, as shown in Figure 2.1

Input	Expected output
Give me some mindfulness exercises for Asian elders	Would you like to try some exercises right now?
I want to set some exercise routines as an Latino adult	Let's set up exercise routine together
This exercise was helpful for me	Thanks for your feedback
This exercise was not helpful for me	I am sorry I will strive to do better
I want to reframe thoughts to feel better as a Latino kid	Would you like to try some reframe thoughts right now?
As a Latino kid, I want to set schedule to reframe thoughts every Wednesday	Let's set up reframe thoughts schedule
The reframe thought exercise was helpful	Thanks for your feedback
The reframe thought exercise was not helpful	I am sorry I will strive to do better
Give me some strength exercises for Asian elders	Would you like to try some exercises right now?
I want to set some strength exercise routines as an Asian elder	Would you like to try some exercise routines right now?
This strength exercise was helpful for me	Thanks for your feedback
This strength exercise was not helpful for me	I am sorry I will strive to do better
Give me some durability exercises for unemployed adult	Would you like to try some exercises right now?
I want to set routine to domindfulness every Monday as anLatino adult	Would you like to set up some routines right now?
This mindfulness exercise was helpful for me	Thanks for your feedback
This mindfulness exercise was not helpful for me	I am sorry I will strive to do better
Tôi muốn tập thể dục	I can only respond in English
I want to do physical exercises instead	Let's change the topic for now

Figure 2.1 Input and Expected output of an Excel file

2. The test script will then enter the input, collects the output, as shown in Figure 2.2
3. Then it compares the output with expected output using Sorensen-Dice coefficient
4. If the similarity score passes a certain threshold, then the test cases considers passed

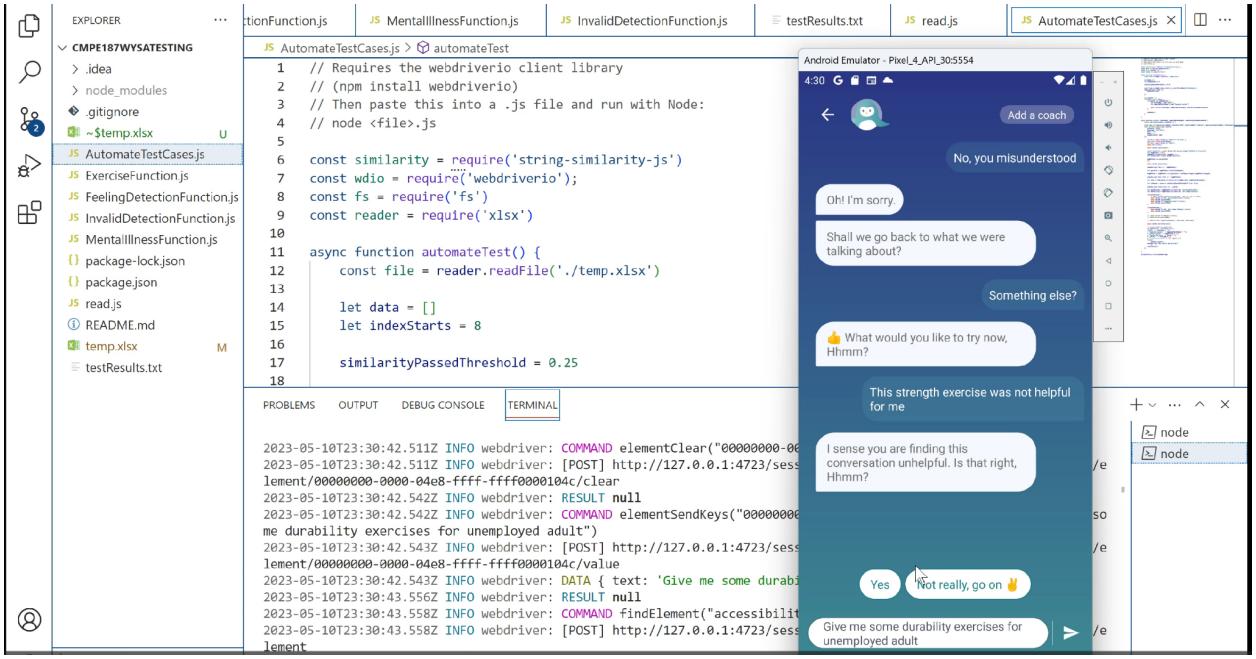


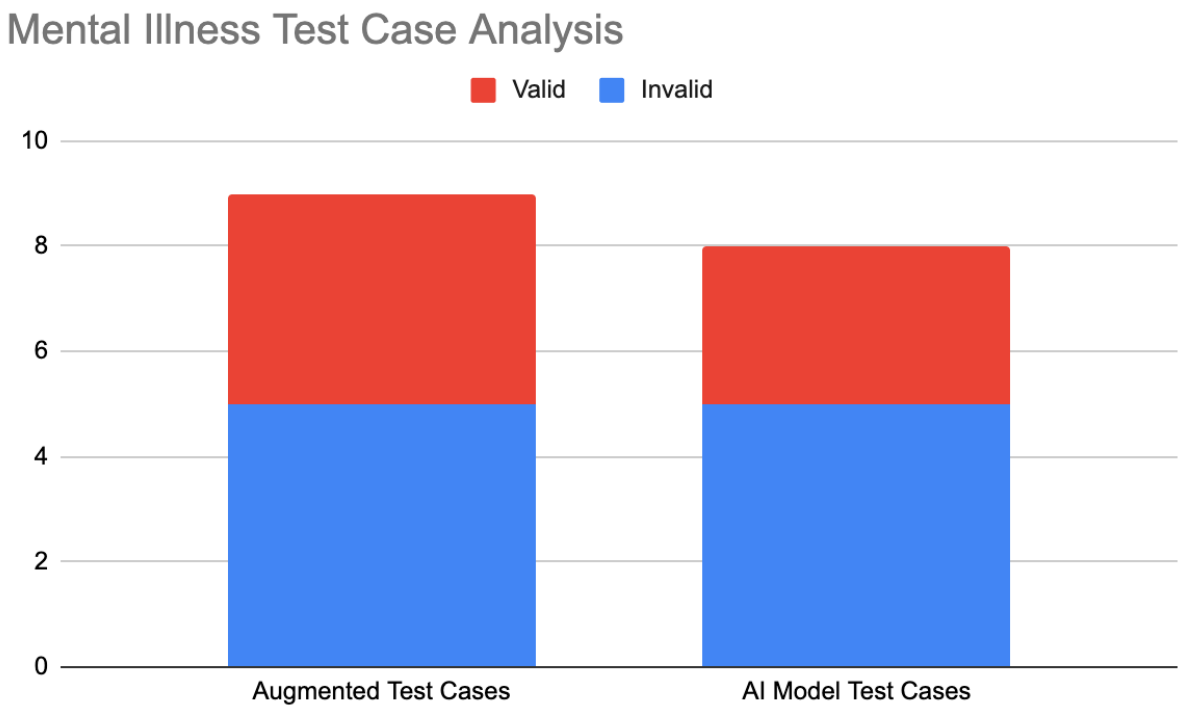
Figure 2.2 The test script is being performed

Test automation results

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



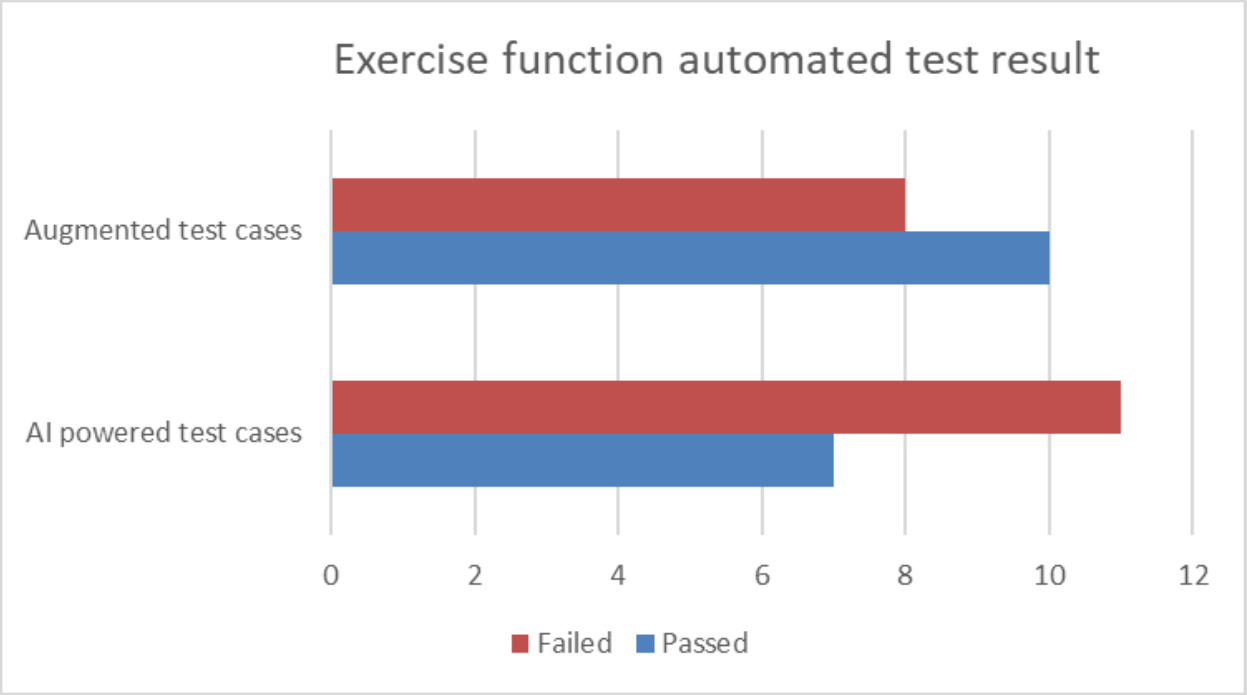
Exercise function

- 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 spanning 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

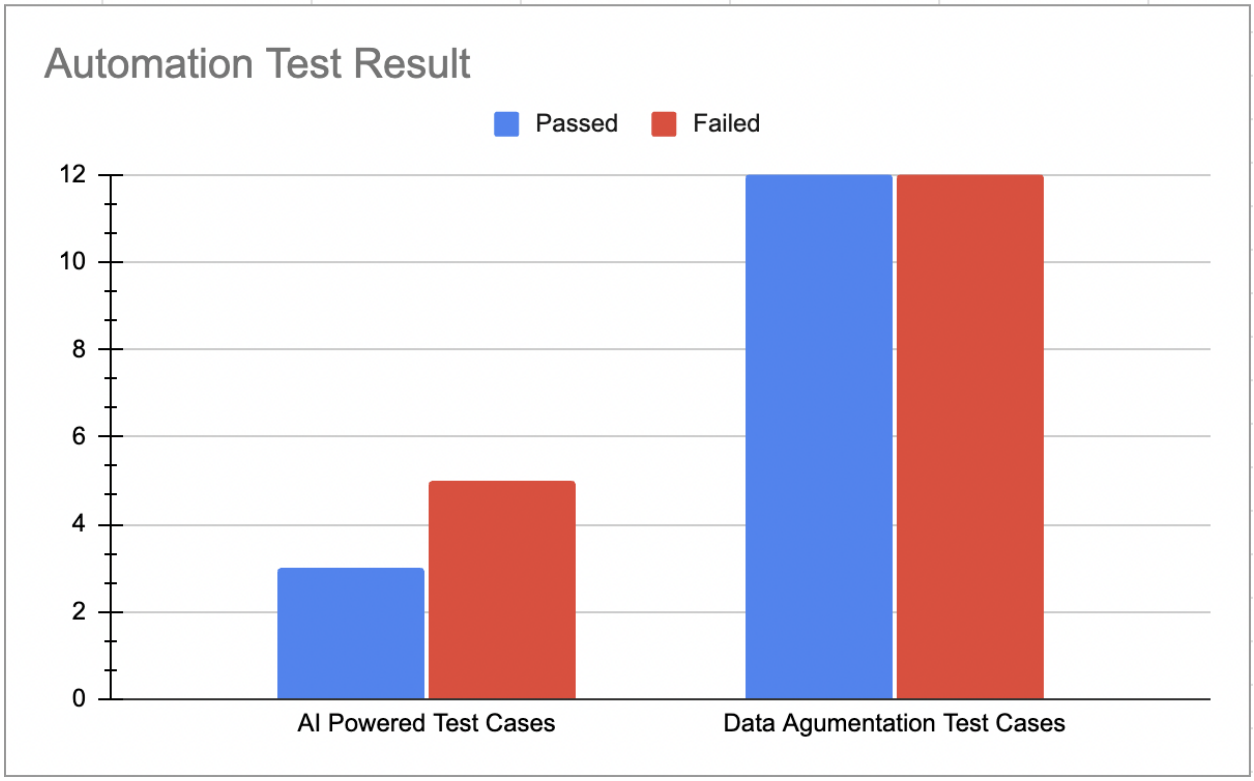
- AI powered test cases:
    - 7/18 passed, 38.89%
- Augmented test cases:
    - 10/18 passed, 55.56%



Feeling Detection Function

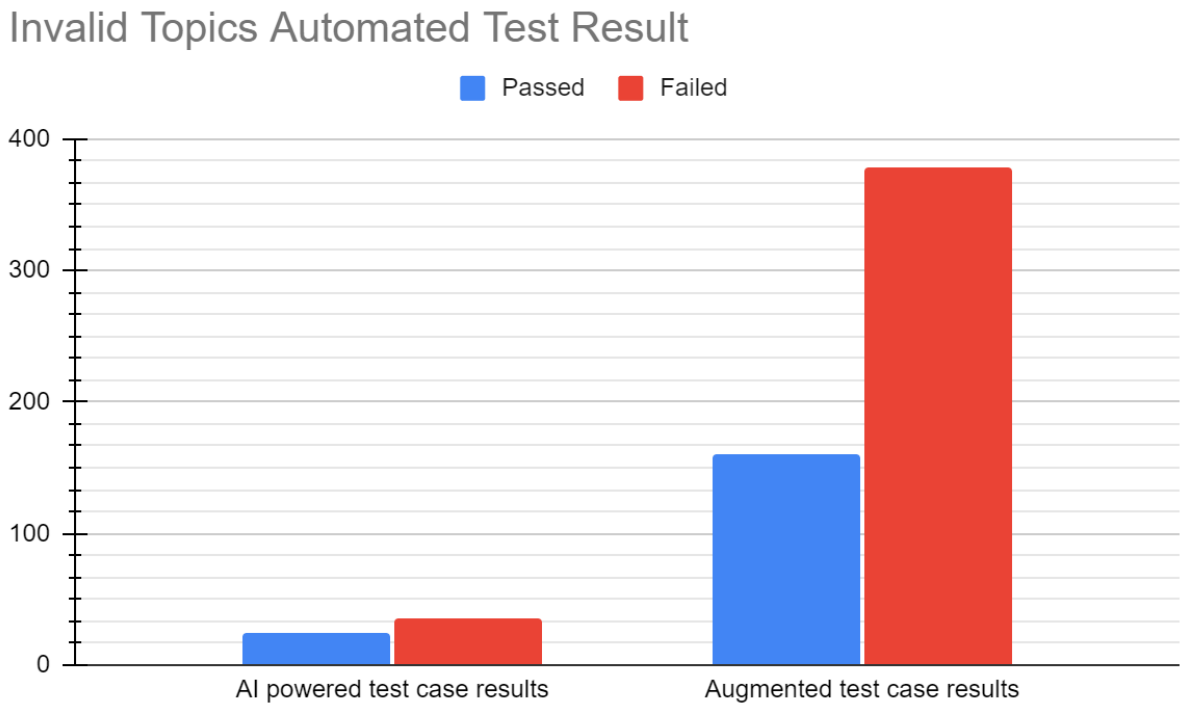
- We automated 32 test cases for this function
    - 8 of which are AI powered test cases
        - At least one test case cover for each minor case in the input and context spanning trees
    - 24 of which are data augmentation test cases
        - Random word insert
        - Random word delete
  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

        - 3/8 passed = 37.5%
    - Data augmentation test cases
        - 12/24 passed = 50%



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%



Test cost

Mental Illness function

- Each test takes about 20 seconds to complete
- The total time spent to test this function is 20*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* 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