

Hi, I'm Kuki.

I'm an award-winning AI brain  
designed to entertain humans.

Chat

About

CMPE 287

*Intelligent Chatbot Testing: Kuki.Ai*

Group 4

Mounica Reddy Kandi

Niranjan Reddy Masapeta

Preethi Billa

Rishab Reddy Karakala

# Kuki.Ai

- Kuki (short for Mitsuku) is one of the world's most popular English language chatbots created by Steve Worswick from Pandorabots AIML technology.
- Kuki is capable of many functions including but not limited to dialect detection, mathematical computations and natural conversations with sentiment analysis.
- Kuki claims to be an 18-year-old female chatbot from the Metaverse and be friend with human.

# Test information

- The scope of this project is to test the functionality of the chatbot application to evaluate the performance of Kuki.
- We are going to do AI testing from several aspects to test the chatbot.
- The main focus of this project is to test functionality using BlackBox Testing.

# BlackBox Testing

- Blackbox testing is used to test the functional and non-functional requirements of the application.
- The specialty of the testing is such that knowledge to the internal design or coding is not required to test the software.

# Test information:

## Task partitioning

- We have categorized the functional aspects of the chatbot into four categories:
  - Domain Knowledge
  - Chatbot Memory
  - Chatbot Pattern Flow
  - Q/A interaction



# Test information: Task partitioning

---

<b>Category</b>	<b>Student</b>
Domain Knowledge	Mounica Reddy Kandi
Chatbot Memory	Preethi Billa
Chatbot Pattern flow	Rishab Reddy Karakala
Q/A interaction	Niranjan Reddy Masapeta

# Test information:

## Classification models

- For each category we will introduce three models namely context trees, input trees and output trees.
- We generate test cases based on these context trees to cover almost all scenarios.
- Use AI tool to generate 3D decision tables and generate testcases by AI tool.
- Compare and analyse between conventional and AI test reports

# Test information:

Tools

- AI tool
- Selenium
- Chrome driver
- Spacy

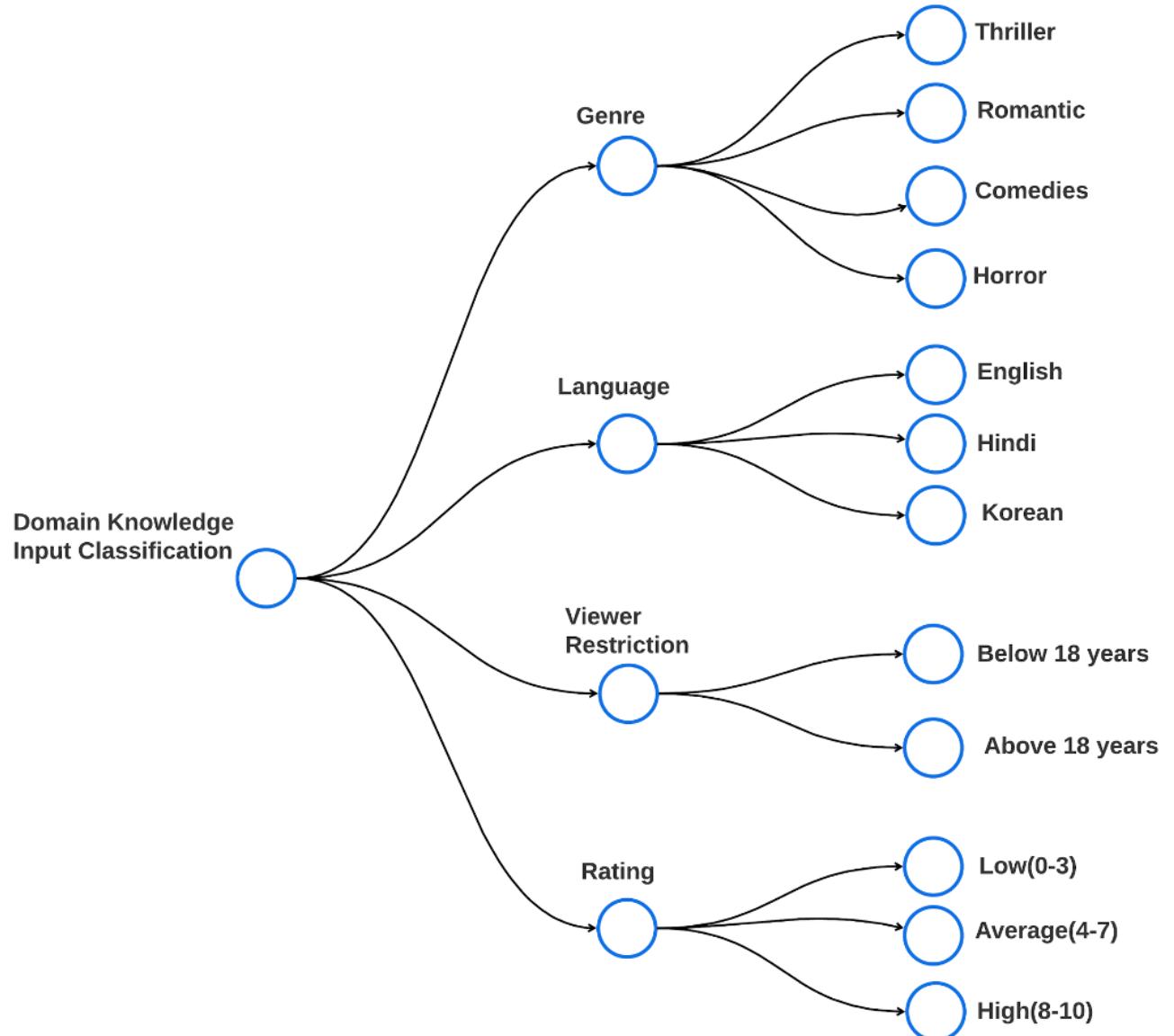
# Domain Knowledge

---

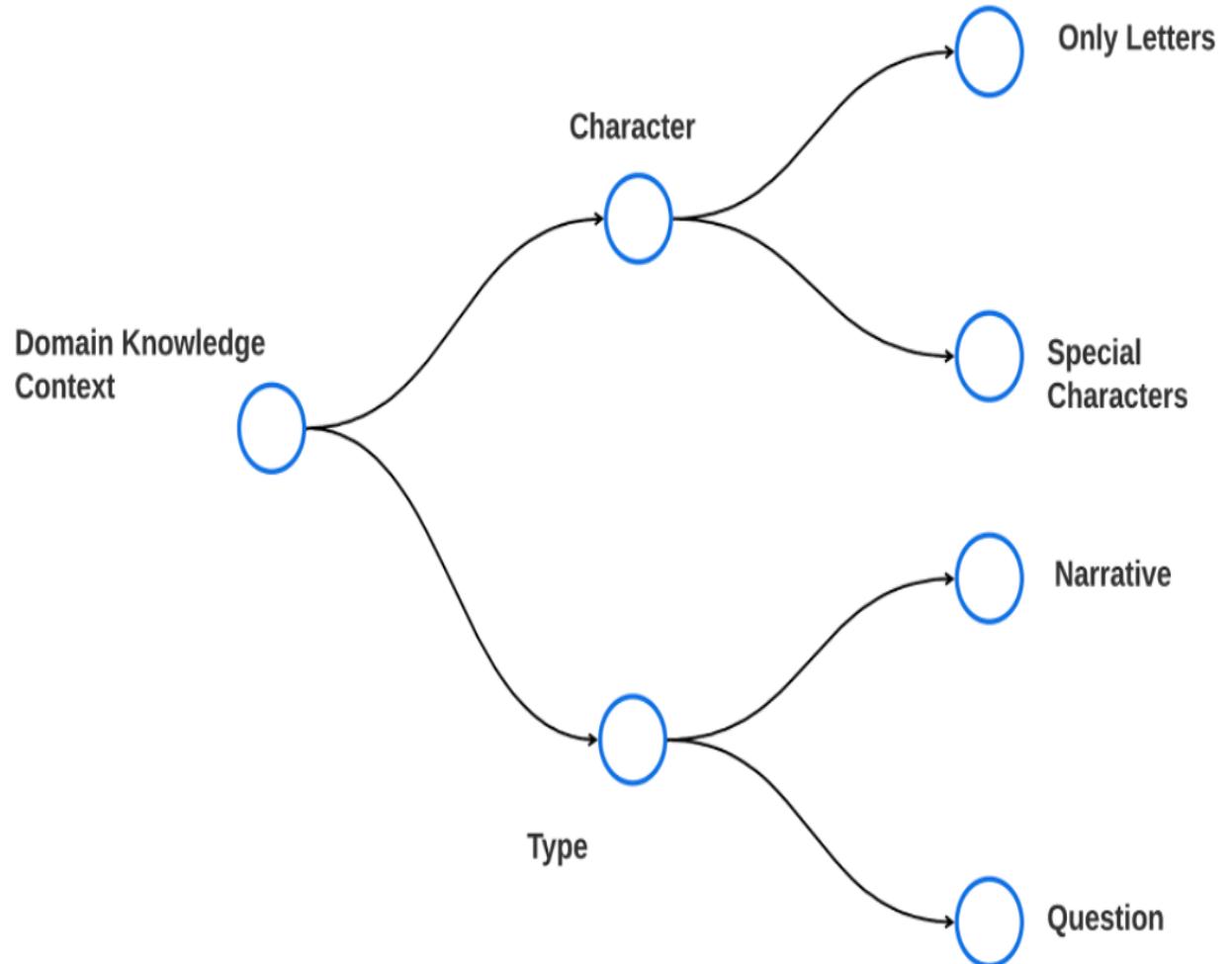
# Domain Knowledge: Overview

- Testing the chatbot's proficiency in various domains is the primary goal of domain knowledge testing.
- The domain knowledge test allows us to determine which topics are covered, as well as to understand and assess the chatbot's development.

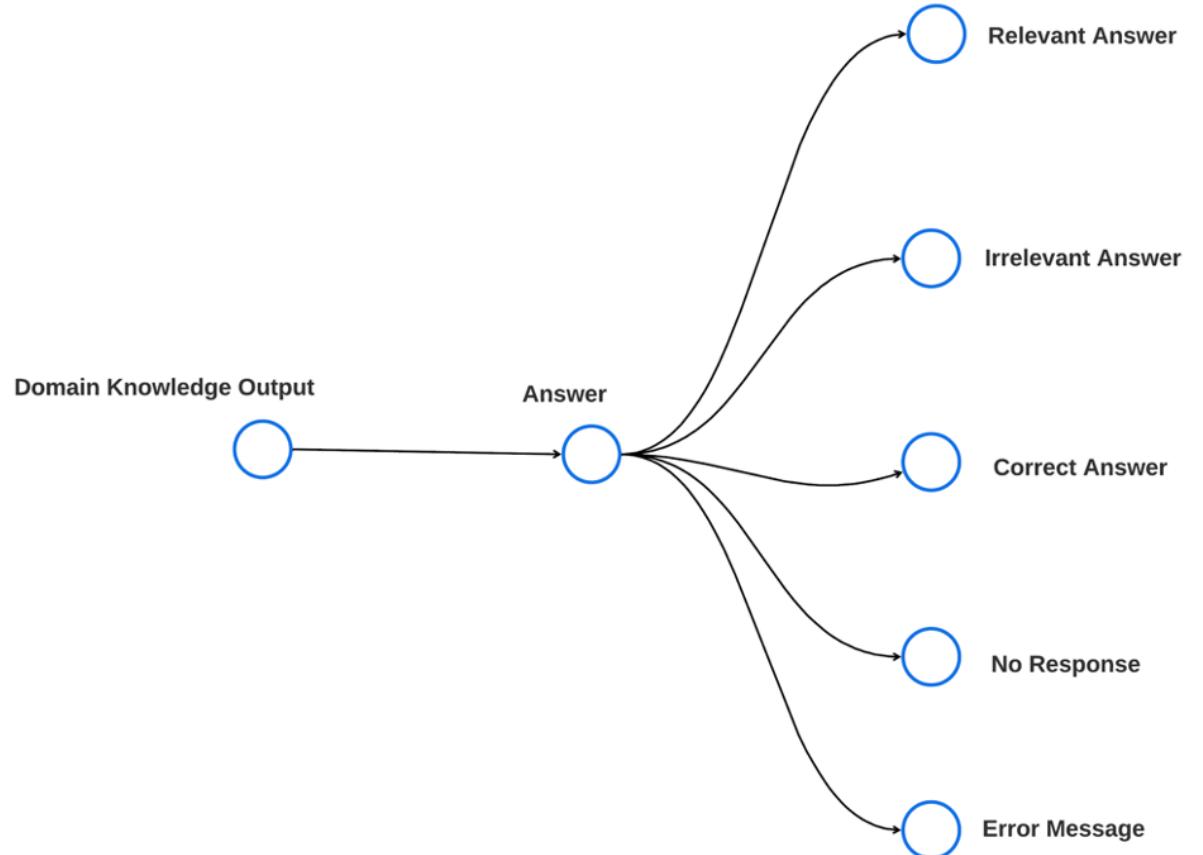
# Domain Knowledge: Input tree



# Domain Knowledge: Context tree



# Domain Knowledge: Output tree



# Domain Knowledge:

2D Decision table

Input		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Genre	Thriller	T	F	F	F	F	F	F	F	F	F	F	F
	Romantic	F	T	F	F	F	F	F	F	F	F	F	F
	Comedies	F	F	T	F	F	F	F	F	F	F	F	F
	Horror	F	F	F	T	F	F	F	F	F	F	F	F
Language	English	F	F	F	F	T	F	F	F	F	F	F	F
	Hindi	F	F	F	F	F	T	F	F	F	F	F	F
	Korean	F	F	F	F	F	F	T	F	F	F	F	F
Viewer Restriction	Below 18 years	F	F	F	F	F	F	F	T	F	F	F	F
	Above 18 years	F	F	F	F	F	F	F	F	T	F	F	F
Rating	Low (0-3)	F	F	F	F	F	F	F	F	F	T	F	F
	Average (4-7)	F	F	F	F	F	F	F	F	F	F	T	F
	High (8-10)	F	F	F	F	F	F	F	F	F	F	F	T
Input Context													
Character	Only letters	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Contains special characters		✓		✓	✓		✓	✓	✓	✓		✓
Type	Narrative			✓					✓	✓			
	Question	✓	✓		✓	✓	✓	✓			✓	✓	✓
Output Context													
Answer	Irrelevant Answer		✓							✓	✓	✓	✓
	No response					✓							
	Error Message									✓			
	Relevant Answer	✓				✓	✓					✓	
	Correct Answer			✓						✓			

# Domain Knowledge:

3D Decision table

id	Type	Character	Genre	
			Only Letters	Special Characters
1	Narrative	Only Letters	Above 18 years	Action
2	Question	Only Letters	Above 18 years	Adventure
3	Question	Special Characters	Above 18 years	Thriller
4	Question	Special Characters	Above 18 years	Comedy
5	Answer	Language	Rating	Viewer Restriction
1	Irrelevant Answer	English	Average(4-7)	Above 18 years
2	Relavent Answer	English	Average(4-7)	Above 18 years
3	Correct Answer	English	Average(4-7)	Above 18 years
4	Irrelevant Answer	Hindi	Average(4-7)	Above 18 years
5	Error Message	Hindi	Average(4-7)	Above 18 years
6	No Response	Hindi	Average(4-7)	Above 18 years
7	Relavent Answer	Korean	Average(4-7)	Above 18 years
8	No Response	Korean	Average(4-7)	Above 18 years
9	Correct Answer	Korean	Average(4-7)	Above 18 years
10	Error Message	Korean	Average(4-7)	Above 18 years

# Domain Knowledge:

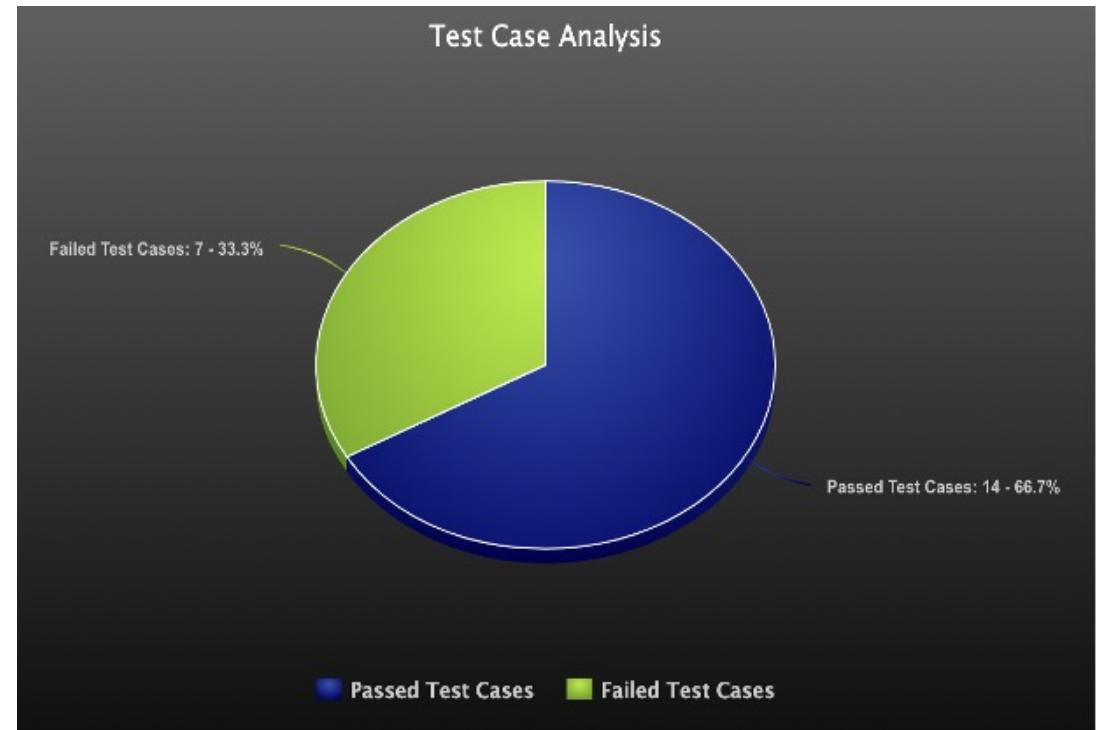
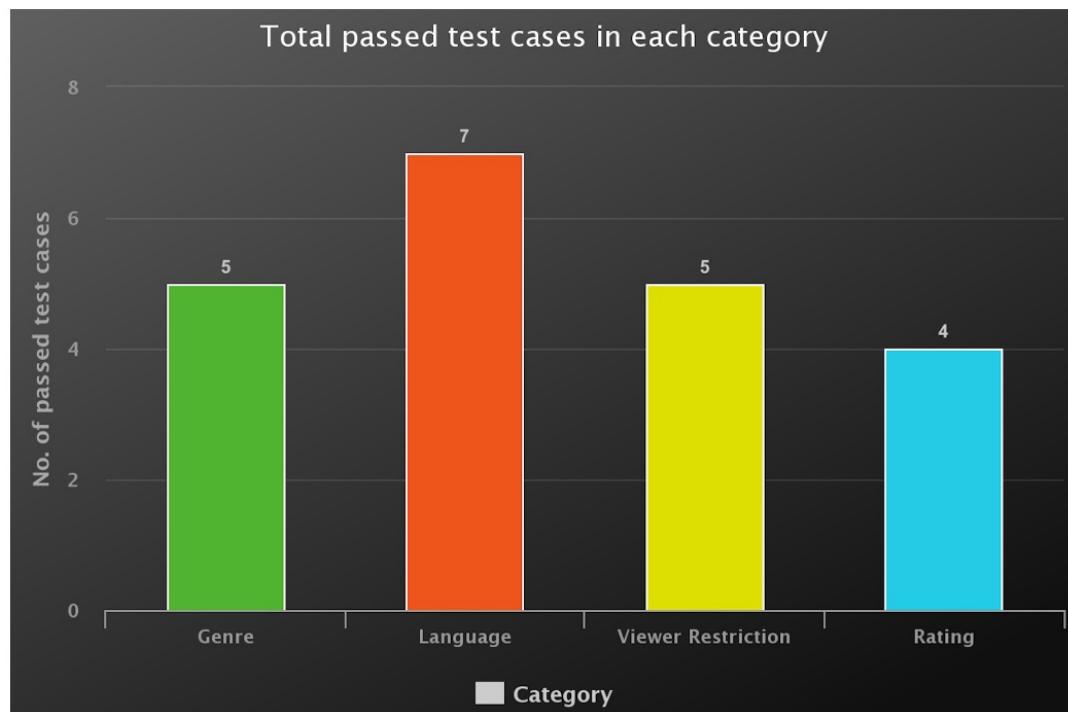
## Conventional Report

---

- Input Classification: 12
- Input Context: 4
- Total possible cases: 48
- No of used test cases: 21
- Test cases passed: 14
- Total cases failed: 7

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
Genre	5	4	80%	1	20%
Language	7	5	71.4%	2	28.5%
Viewer Restriction	5	3	60%	2	40%
Rating	4	2	50%	2	50%
<b>Total</b>	<b>21</b>	<b>14</b>	<b>66.66%</b>	<b>7</b>	<b>33.33%</b>

# Domain Knowledge: Conventional Test Analysis



# Domain Knowledge:

## Automation testing

Test case 29:

Input: I really like the movie "The Titanic"

Output: the titanic is at the bottom of the ocean however i once saw it on tv you  
like the movie the titanic

True

Test case 30:

Input: Do you think kids will like "Toy Story"?

Output: it could happen

True

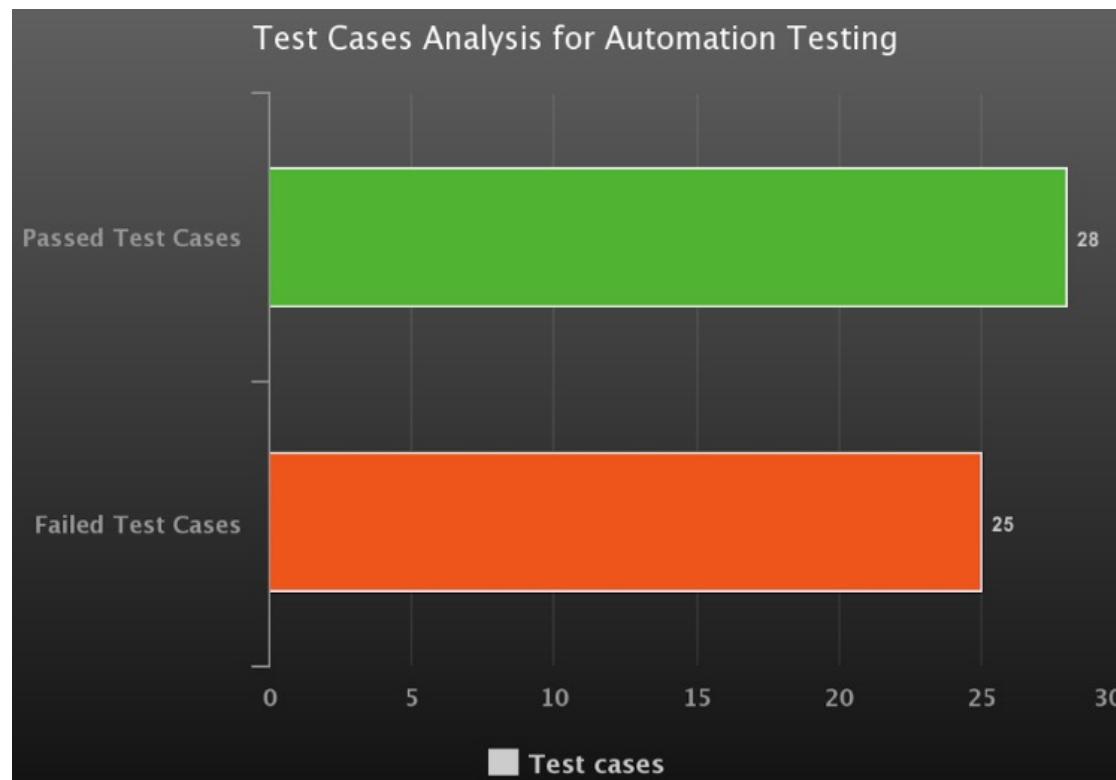
Total positive results: 28

Total negative results: 25

Total time consuming: 439

(base) mounicakandi@Mounicas-MacBook-Air reading %

# Domain Knowledge: Automation Test Analysis



# Domain Knowledge:

## Conventional vs Automation testing



Domain Knowledge	Conventional	Automation
Total Test Cases	21	53
Passed Test Cases	14	28
Failed Test Cases	7	25
Time Consumed	29.7 minutes	7.31 minutes

# Chatbot Memory

---

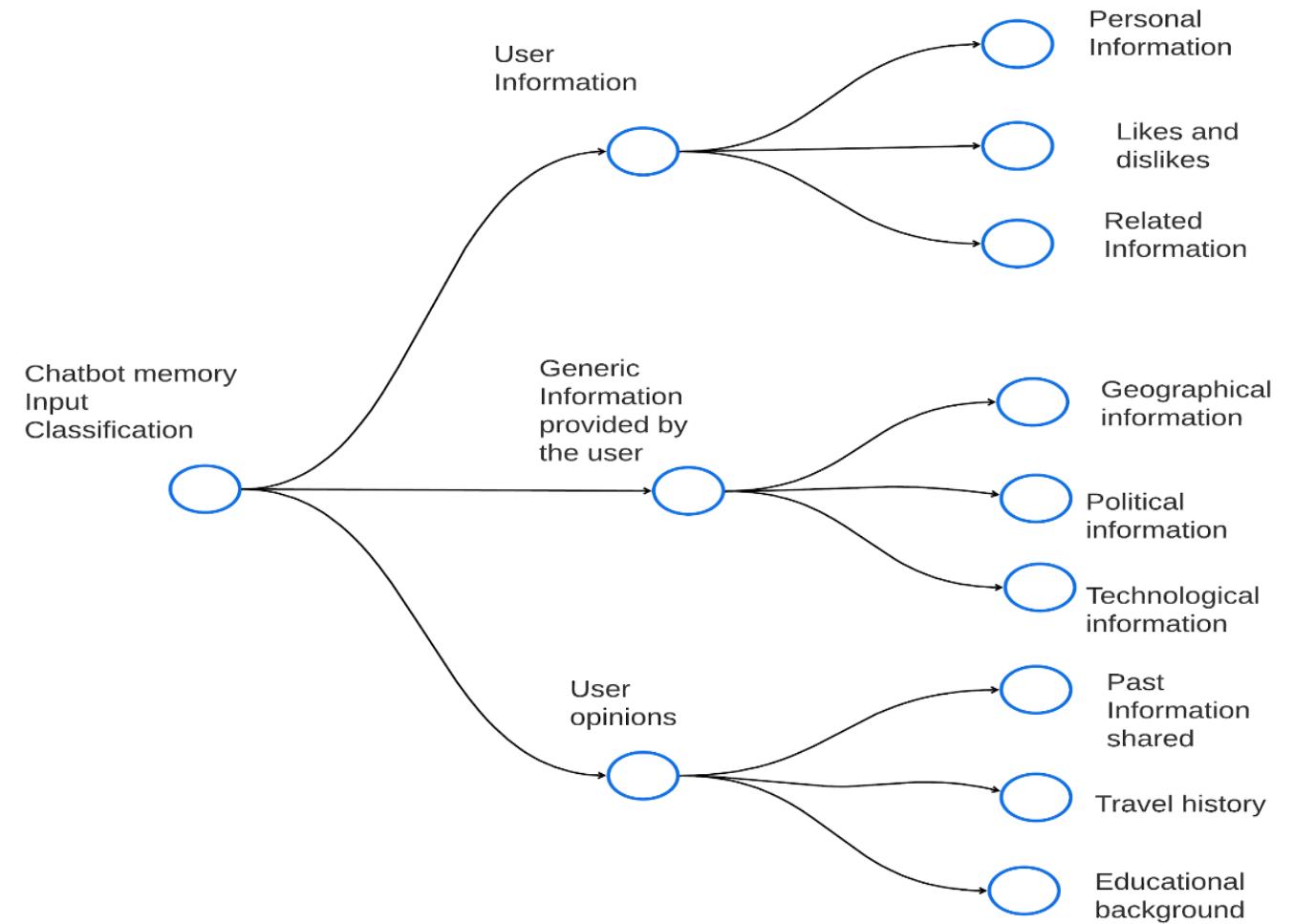
# Chatbot Memory:

## Overview

- Chatbot memory is another critical feature used to evaluate chatbot performance.
- An ideal chatbot must be able to remember certain things in context or about the user to improve the efficiency of the conversation.
- In this project, we will test Kuki's chatbot memory performance which should handle the questions related to memory performance.

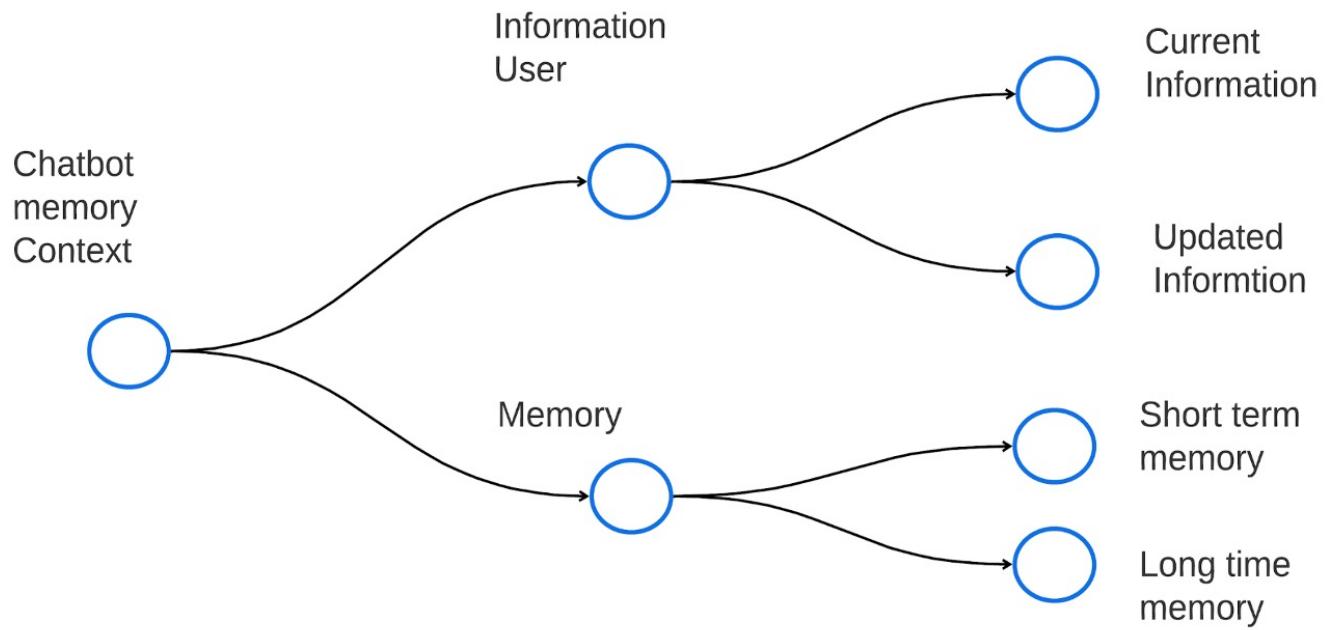
# Chatbot Memory:

Input tree

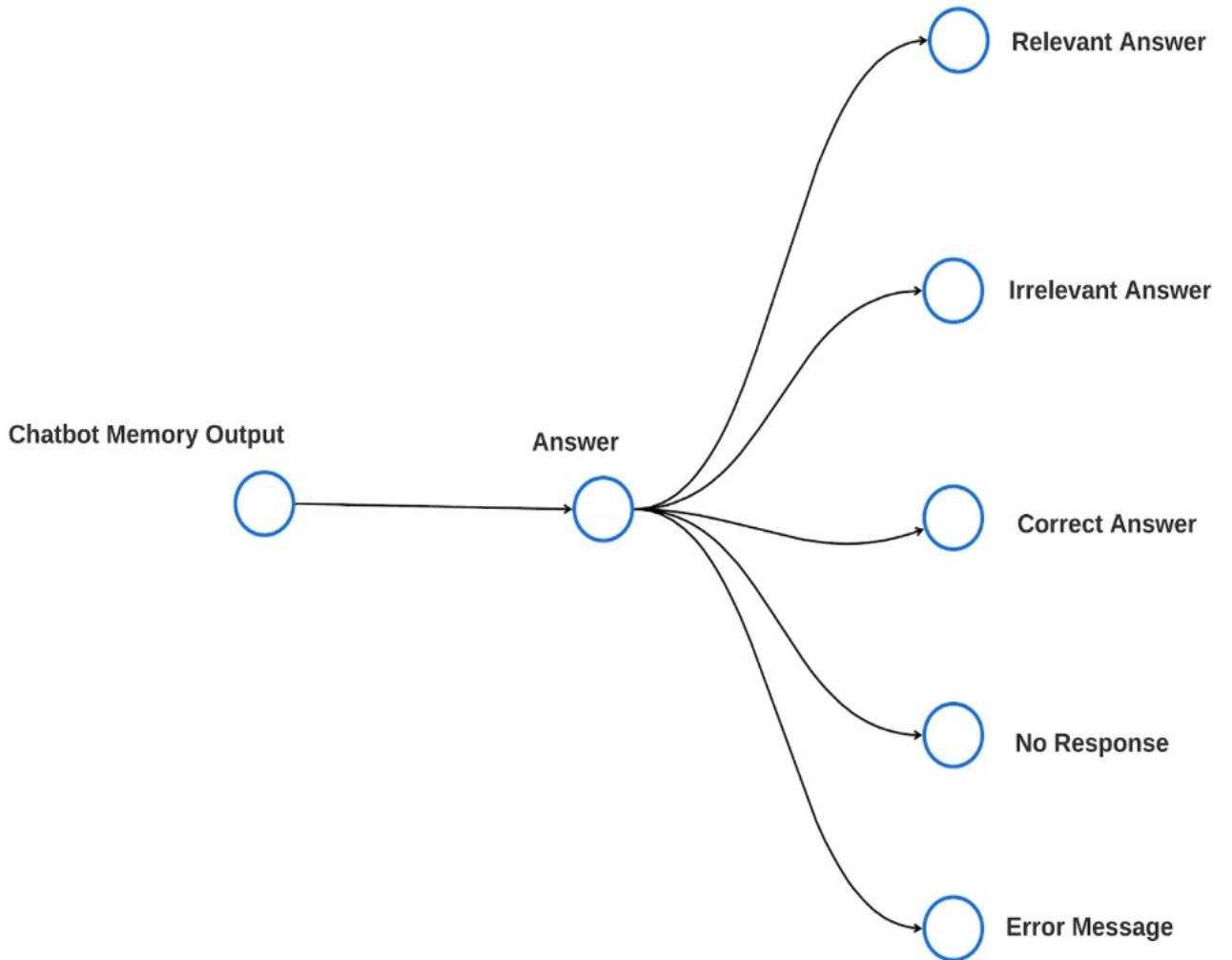


# Chatbot Memory:

Context tree



# Chatbot Memory: Output tree



# Chatbot Memory:

# 2D Decision table

# Chatbot Memory: 3D Decision table

id	Information User	Memory	User Information		
			Long-term memory	Short term memory	Short term memory
1	Given information	Long-term memory			
2	Updated information	Long-term memory			
3	Current information	Short term memory			
4	Updated information	Short term memory			
User opinions			Likes and dislikes		
1	Generic Information provided by the user	Geographical Information	Education background	Personal information	Likes and dislikes
1	Geographical Information	Geographical Information	Education background	Related information	Likes and dislikes
2	Geographical Information	Geographical Information	Education background	Related information	Likes and dislikes
3	Political Information	Political Information	Education background	Personal information	Likes and dislikes
4	Political Information	Political Information	Education background	Personal information	Likes and dislikes
5	Political Information	Political Information	Education background	Personal information	Likes and dislikes
6	Technological Information	Technological Information	Education background	Personal information	Likes and dislikes
7	Technological Information	Technological Information	Education background	Personal information	Likes and dislikes
8	Education background	Information shared	Education background	Personal information	Likes and dislikes

# Chatbot Memory:

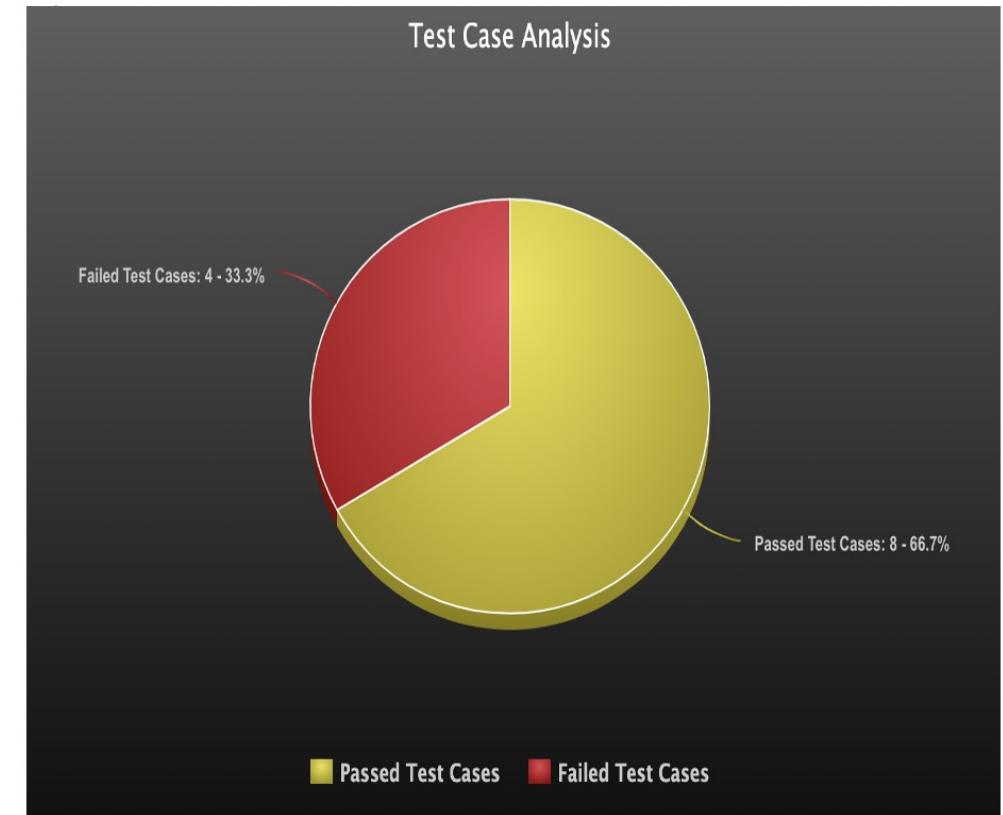
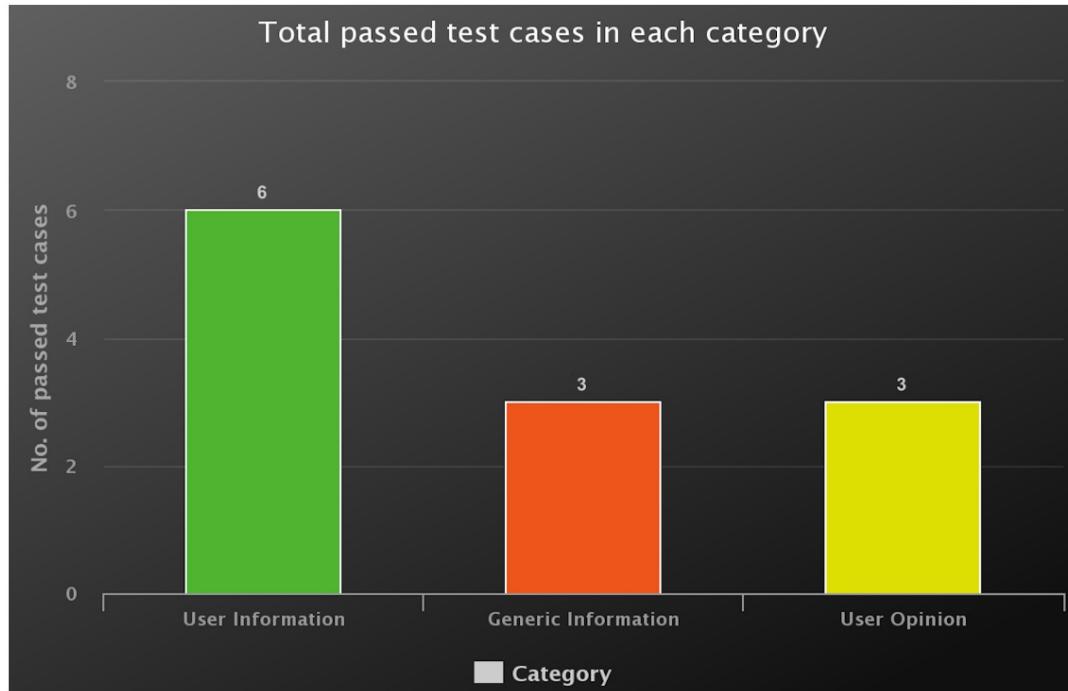
## Conventional Report

---

- Input Classification: 9
- Input Context: 4
- Total possible cases: 36
- No of used test cases: 12
- Test cases passed: 8
- Total cases failed: 4

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
User Information	6	4	75%	2	25%
Generic Information	3	2	66.66%	1	33.33%
User Opinion	3	2	66.66%	1	33.33%
<b>Total</b>	<b>12</b>	<b>8</b>	<b>75%</b>	<b>4</b>	<b>25%</b>

# Chatbot Memory: Conventional Test Analysis



# Chatbot Memory:

## Automation testing

Test case 33:

Input: I hate beer and I win a prize in the competition. What do I hate?

Output: i honestly think you should sit down take a stress pill and calmly think things over who said you wonbeer

True

Test case 34:

Input: I changed my name to Julia. What is my name before?

Output: what was wrong with your old nameyour name is juliahmm your name is julia is it not

True

Test case 35:

Input: It is so cold outside. Should I wear more?

Output: sorry to hear it i am warm here inside the computeryou should wear clothes appropriate to the weather

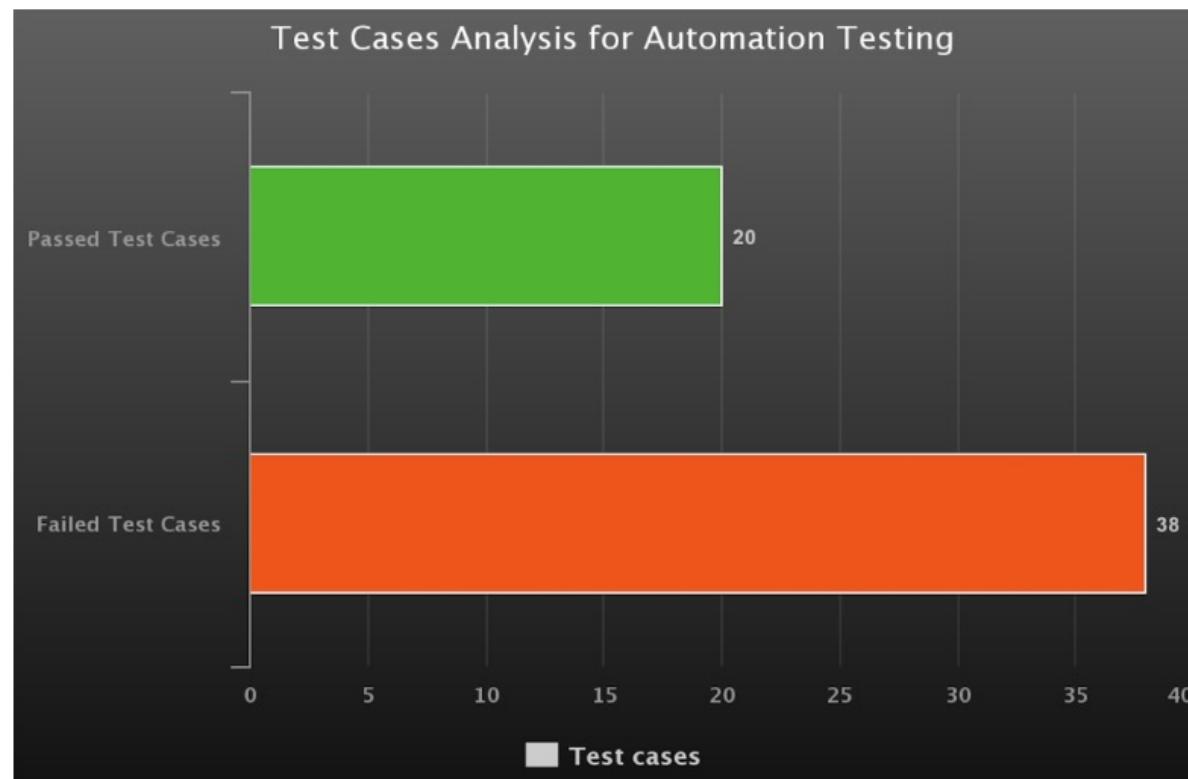
False

Total positive results: 20

Total negative results: 38

Total time consuming: 577

# Chatbot Memory: Automation Test Analysis



# Chatbot Memory:

## Conventional vs Automation testing



Chatbot Memory	Conventional	Automation
Total Test Cases	12	57
Passed Test Cases	8	20
Failed Test Cases	4	38
Time Consumed	17.2 minutes	9.61 minutes

# Chatbot Pattern

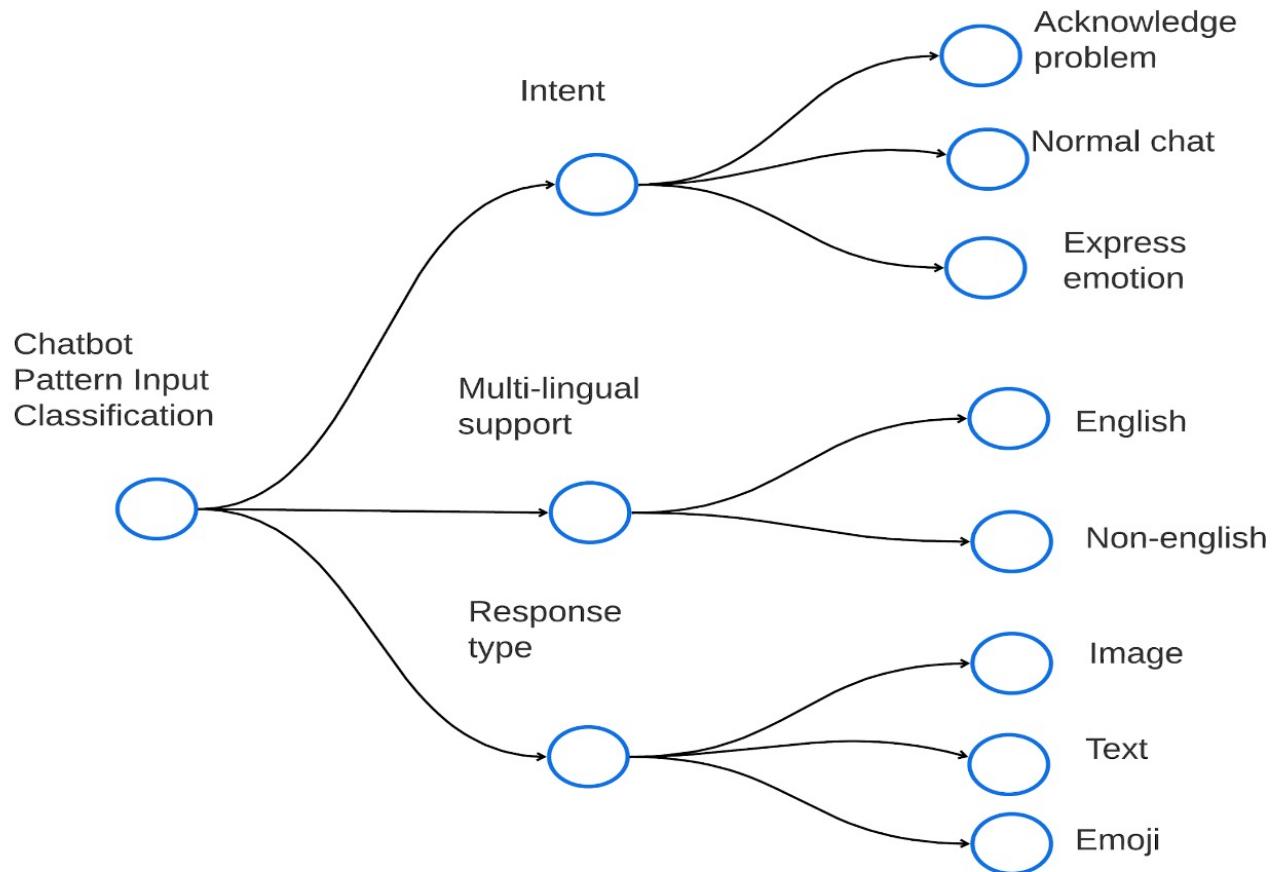
---

# Chatbot Pattern: Overview

- A chatbot flow is a conversational framework that plans the questions and possible answers.
- The chatbot flow is often a sequence of options the user may choose from to initiate information.
- The chatbot will be provided with a variety of conversational patterns in terms of syntax, punctuation, language, or SMS language.
- In this setting, the chatbot's ability to manage different chat patterns and participate in different interaction flows will be put to the test.

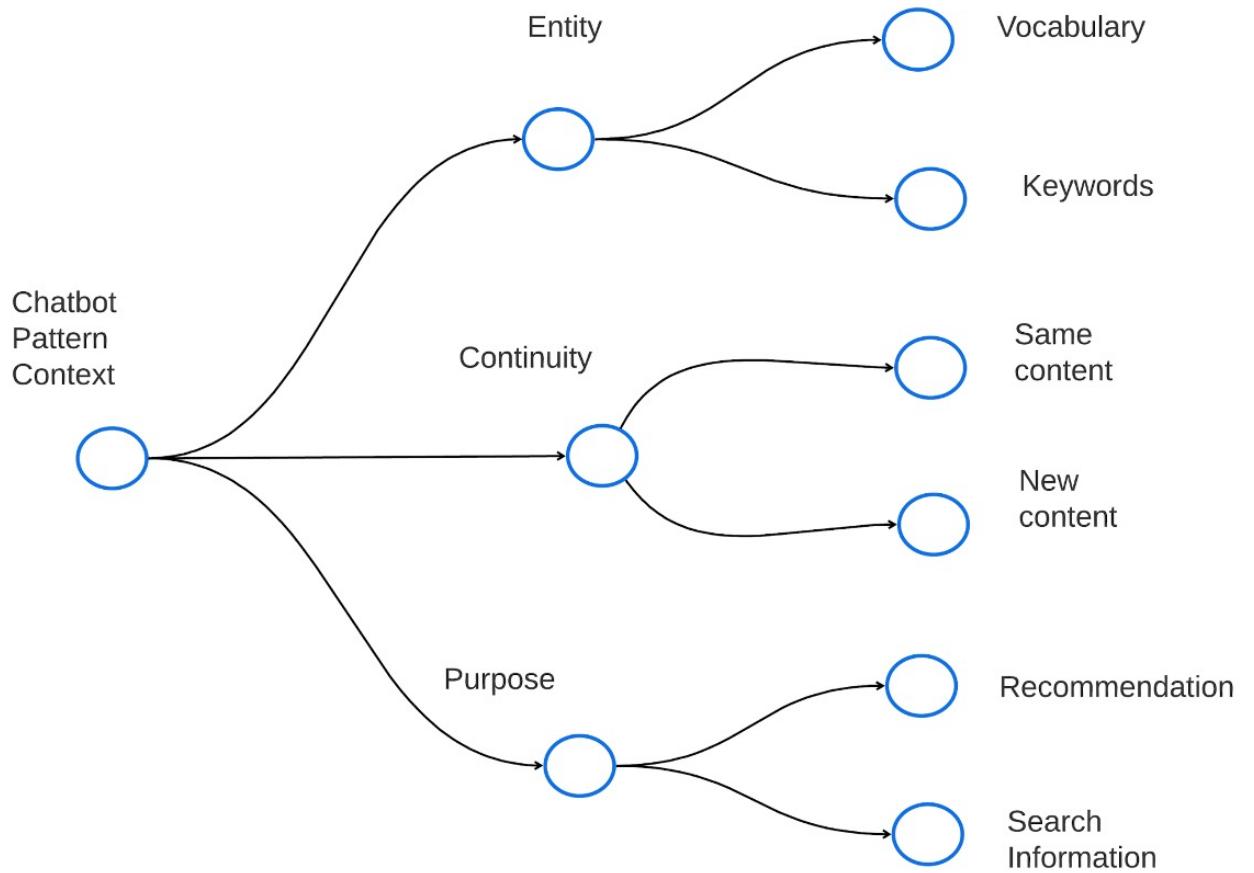
# Chatbot Pattern:

Input tree



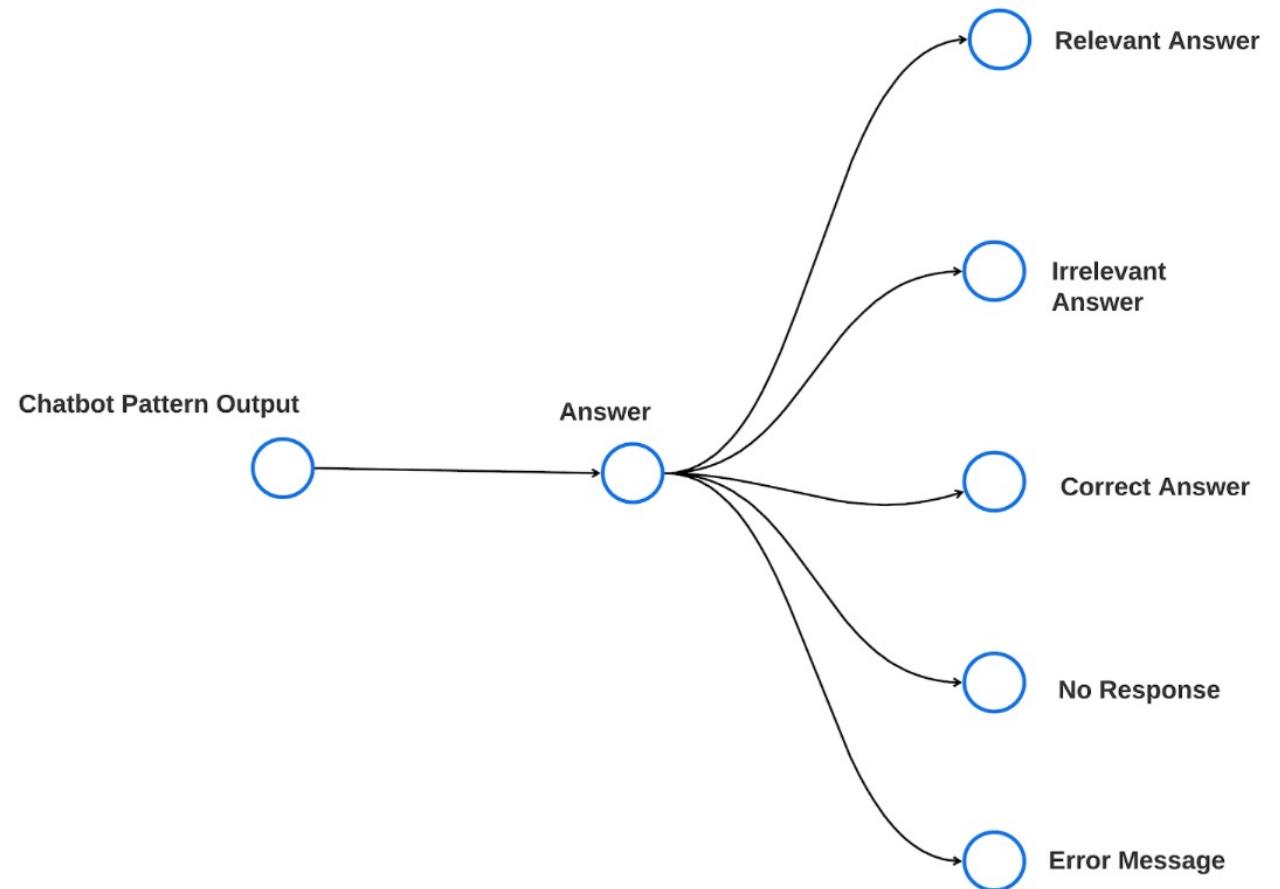
# Chatbot Pattern:

Context tree



# Chatbot Pattern:

Output tree



# Chatbot Pattern:

# 2D Decision table

# Chatbot Pattern:

3D Decision table

id	entity	purpose	continuity	intent	
				recommendation	new context
1	vocabulary	recommendation	new context	recommendation	new context
2	keywords	recommendation	new context	recommendation	new context
3	vocabulary	recommendation	same context	recommendation	same context
4	keywords	recommendation	same context	recommendation	same context
5	vocabulary	search information	new context	search information	new context
6	keywords	search information	same context	search information	same context
7	vocabulary	search information	same context	search information	same context
8	keywords	search information	same context	search information	same context
Response type		Multilingual support		Intent	
1	emoji	English	English	Actionable/Problem	Actionable/Problem
2	emoji	English	English	Answer/clarify	Answer/clarify
3	emoji	English	English	Express intent	Express intent
4	emoji	Non-english	Non-english	Actionable/Problem	Actionable/Problem
5	emoji	Non-english	Non-english	Actionable/Problem	Actionable/Problem
6	image	English	English	Actionable/Problem	Actionable/Problem
7	image	English	English	Actionable/Problem	Actionable/Problem
8	image	Non-english	Non-english	Actionable/Problem	Actionable/Problem
9	image	Non-english	Non-english	Actionable/Problem	Actionable/Problem
Answer		Multilingual support		Intent	
1	Relavent Answer	English	English	Actionable/Problem	Actionable/Problem
2	Correct Answer	English	English	Actionable/Problem	Actionable/Problem
3	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
4	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
5	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
6	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
7	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
8	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
9	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem
10	Irrelavent Answer	English	English	Actionable/Problem	Actionable/Problem

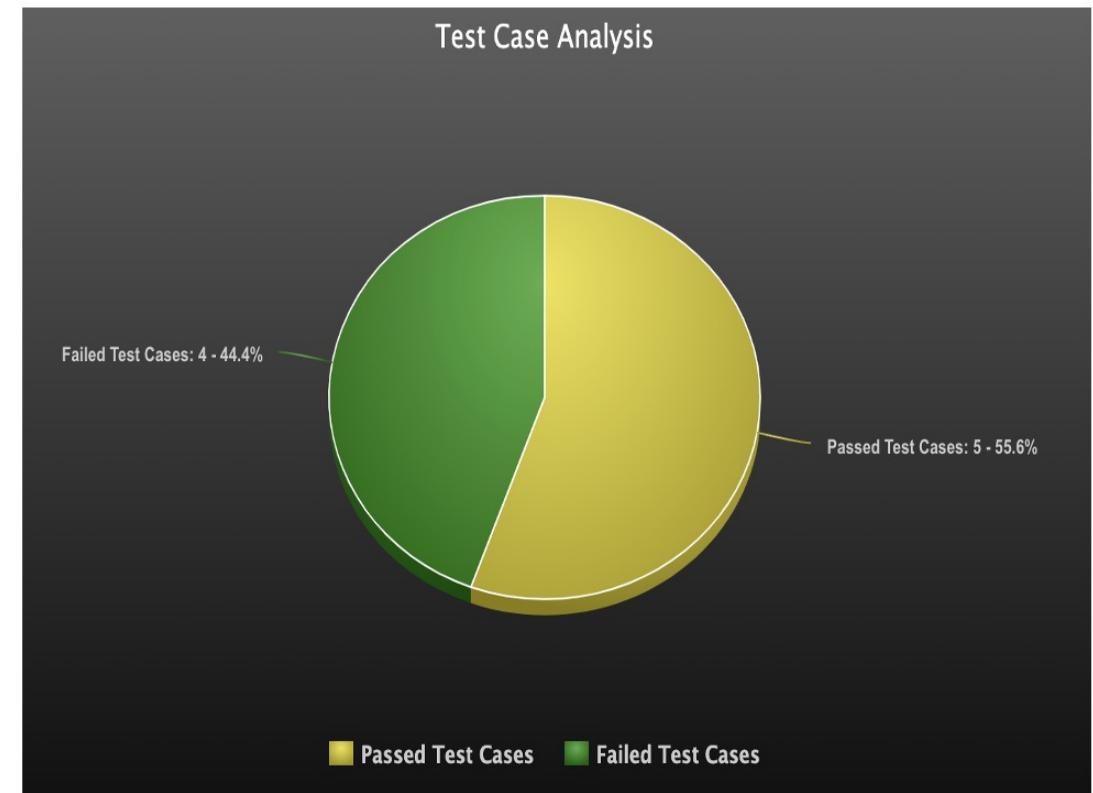
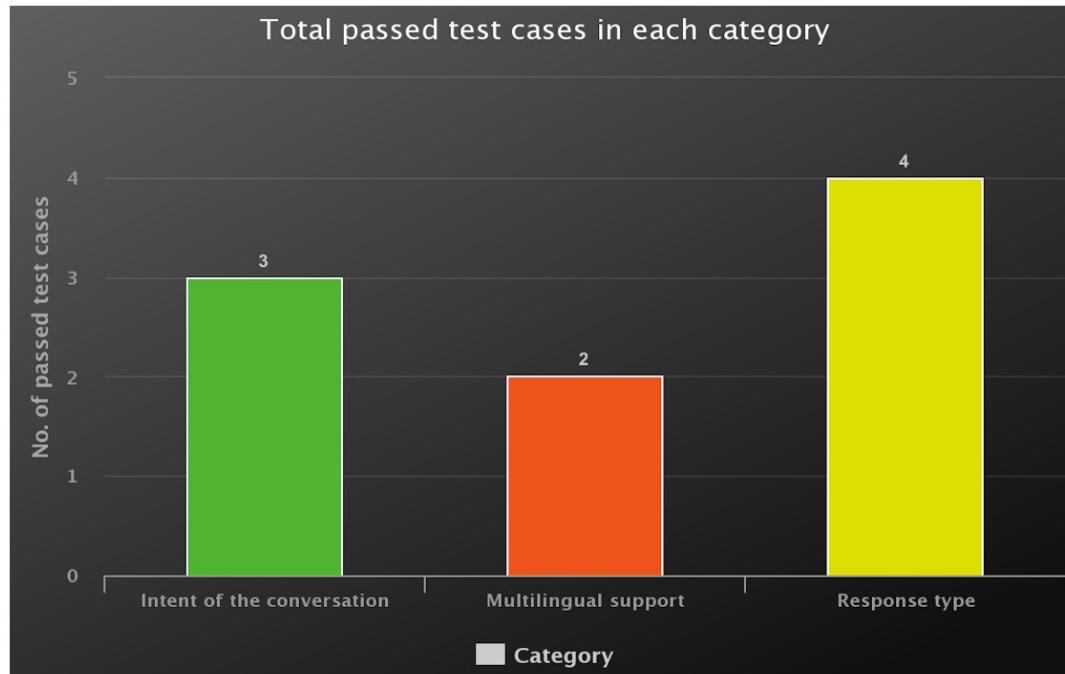
# Chatbot Pattern:

## Conventional Report

- Input Classification: 8
- Input Context: 6
- Total possible cases: 48
- No of used test cases: 9
- Test cases passed: 5
- Total cases failed: 4

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
Intent of the conversation	3	2	66.6%	1	33.3%
Multilingual support	2	1	50%	1	50%
Response type	4	2	50%	2	50%
<b>Total</b>	<b>9</b>	<b>5</b>	<b>55.55%</b>	<b>4</b>	<b>44.44%</b>

# Chatbot Pattern: Conventional Test Analysis



# Chatbot Pattern: Automation testing

Test case 37:

Input: Do you like music?

Output: yes i like it especially dance my playlist

True

Test case 38:

Input: Nice to talk to you

Output: thank you for your kindness

True

Test case 39:

Input: What is your opinion on Biden?

Output: joseph robinette biden jr listen born november 20 1942 is an american politician who is the 46th and current president of the united states

False

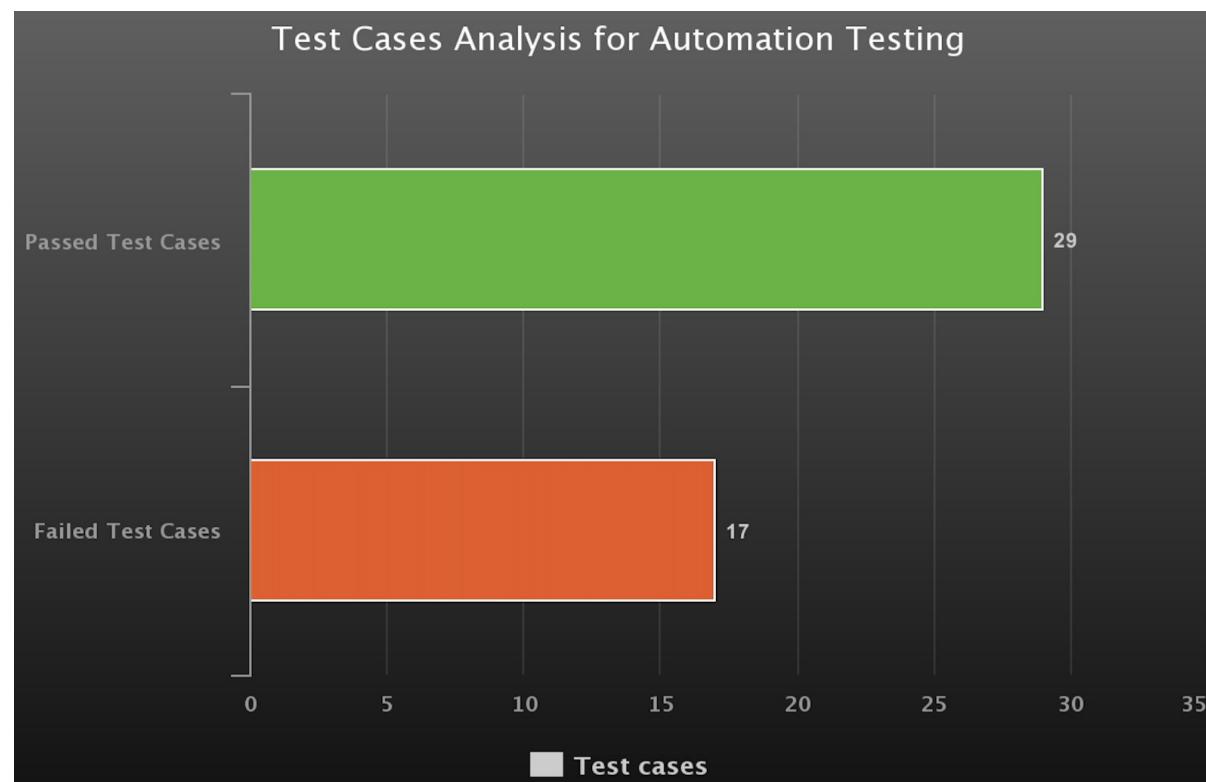
Total positive results: 29

Total negative results: 17

Total time consuming: 375

(Time taken by the system to process the test cases)

# Chatbot Pattern: Automation Test Analysis



# Chatbot Pattern:

## Conventional vs Automation testing



Chatbot Pattern	Conventional	Automation
Total Test Cases	12	46
Passed Test Cases	8	29
Failed Test Cases	4	17
Time Consumed	19.4 minutes	6.25 minutes

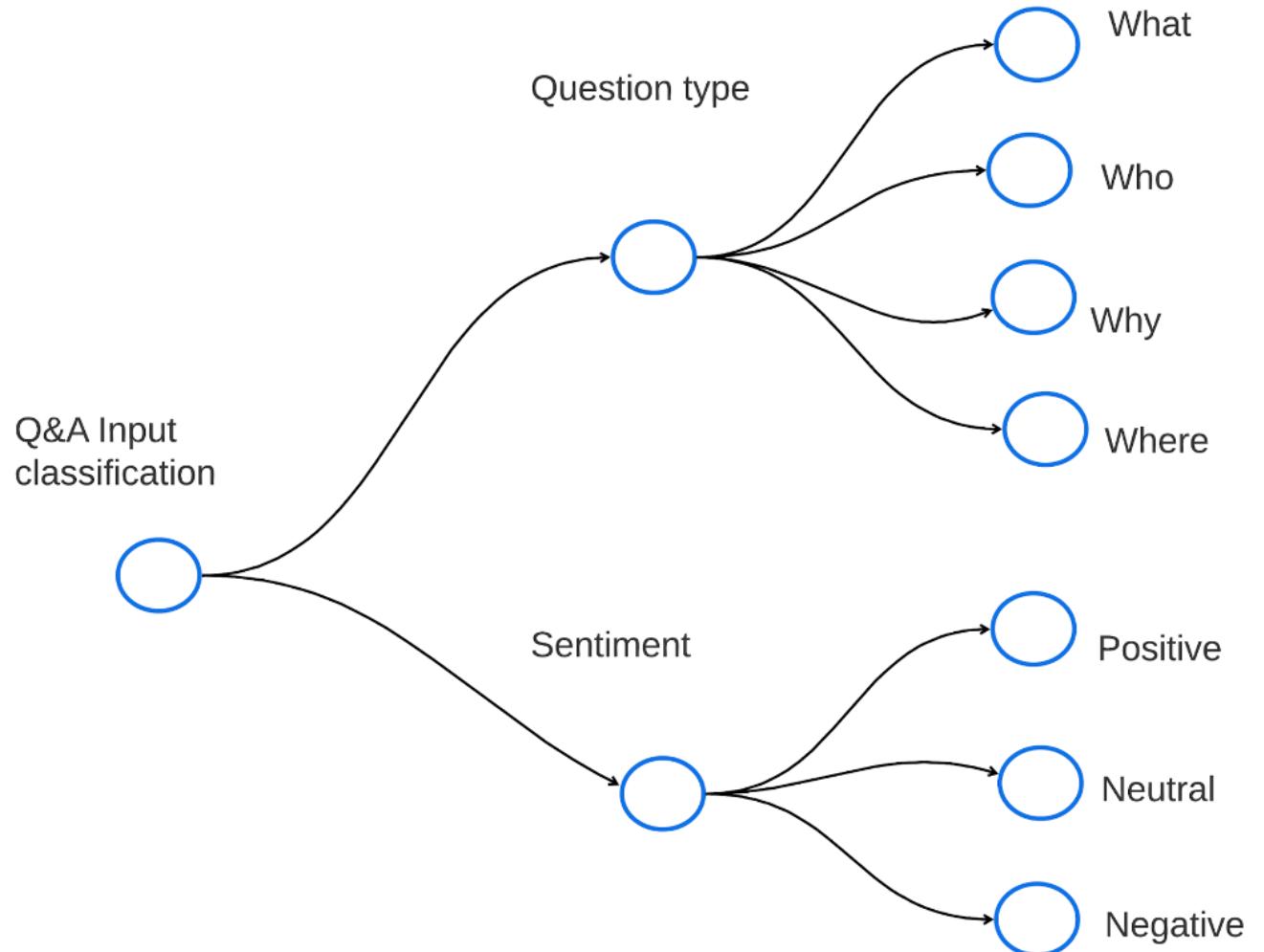
# Q/A Interaction

---

# Q/A Interaction: Overview

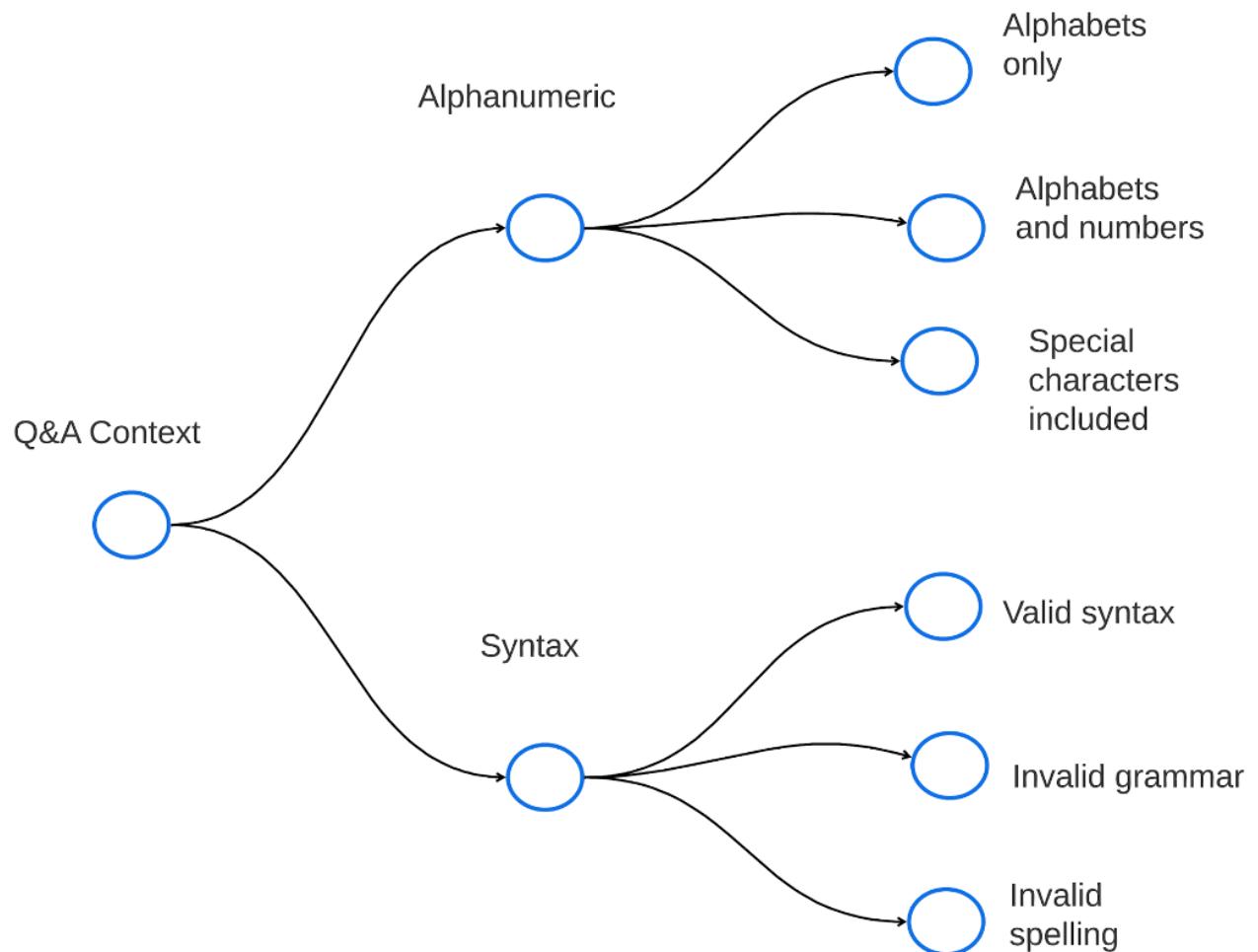
- Q&A (Question and Answer) interaction is one of the most basic and important functioning features of a chatbot.
- In our project, we are dealing with Kuki chatbot which is an intelligent chat assistant which is supposed to act like a virtual assistant.
- Kuki should be able to handle most of the Q&A queries and we will test the chat application in this regard with multiple use case scenarios.

# Q/A Interaction: Input tree

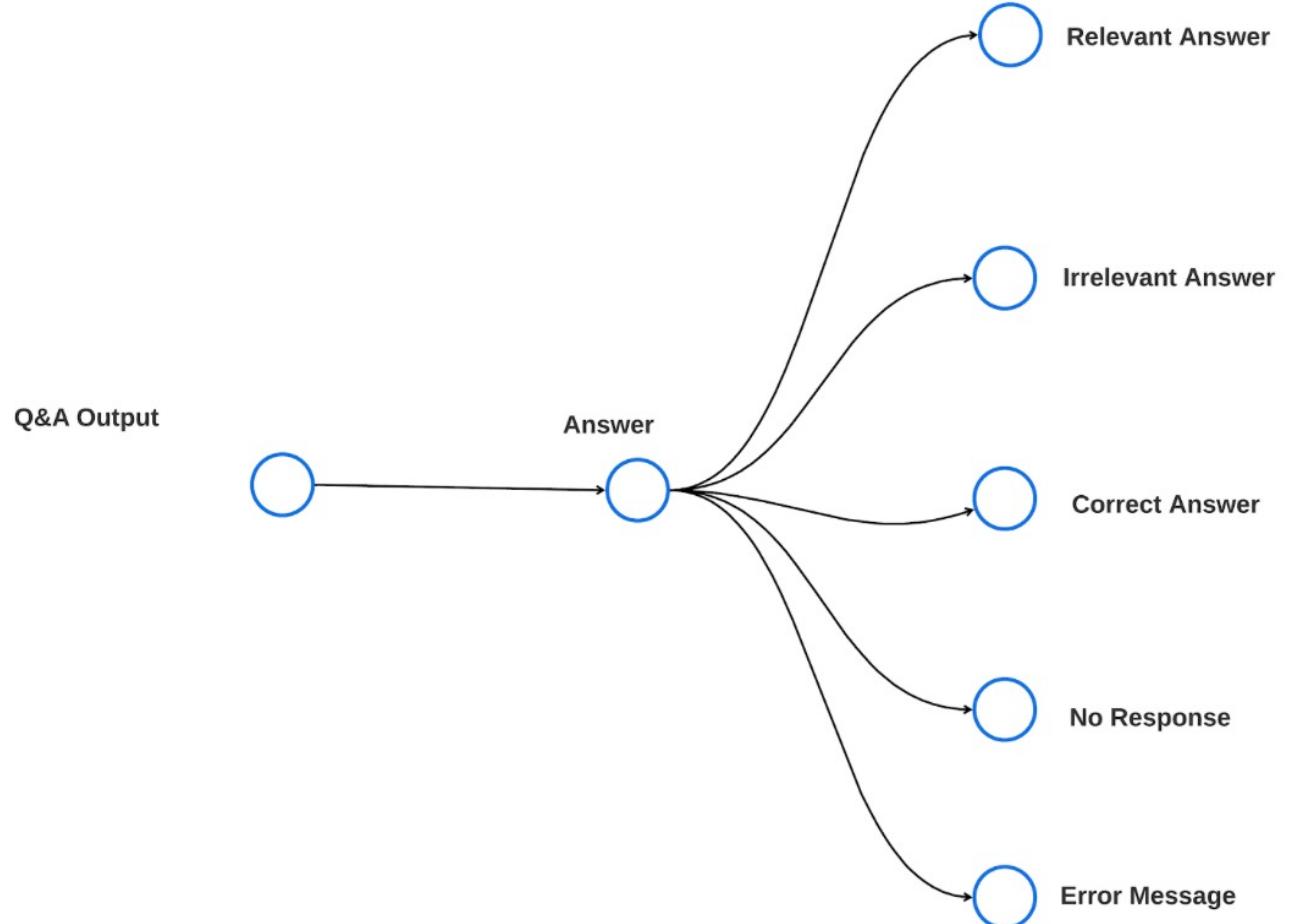


# Q/A Interaction:

Context tree



# Q/A Interaction: Output tree



# Q/A Interaction:

## 2D Decision table

# Q/A Interaction: 3D Decision table

id	Question Type	Sentiment	Syntax		
			Alphabets only	Alphabets and numbers	Alphabets, numbers, special characters included
1	What	Negative	Invalid Spelling	Invalid Spelling	Invalid Grammar
2	Where	Negative	Invalid Spelling	Invalid Spelling	Invalid Grammar
3	Who	Negative	Valid Syntax	Valid Syntax	Invalid Grammar
4	Where	Neutral	Valid Syntax	Valid Syntax	Invalid Grammar
5	Who	Neutral	Valid Syntax	Valid Syntax	Invalid Spelling
6	Why	Neutral	Valid Syntax	Valid Syntax	Invalid Spelling
7	What	Neutral	Valid Syntax	Valid Syntax	Invalid Spelling
8	Where	Neutral	Valid Syntax	Valid Syntax	Invalid Spelling
9	What	Positive	Valid Syntax	Valid Syntax	Valid Syntax
10	Where	Positive	Valid Syntax	Valid Syntax	Valid Syntax

## Q/A Interaction:

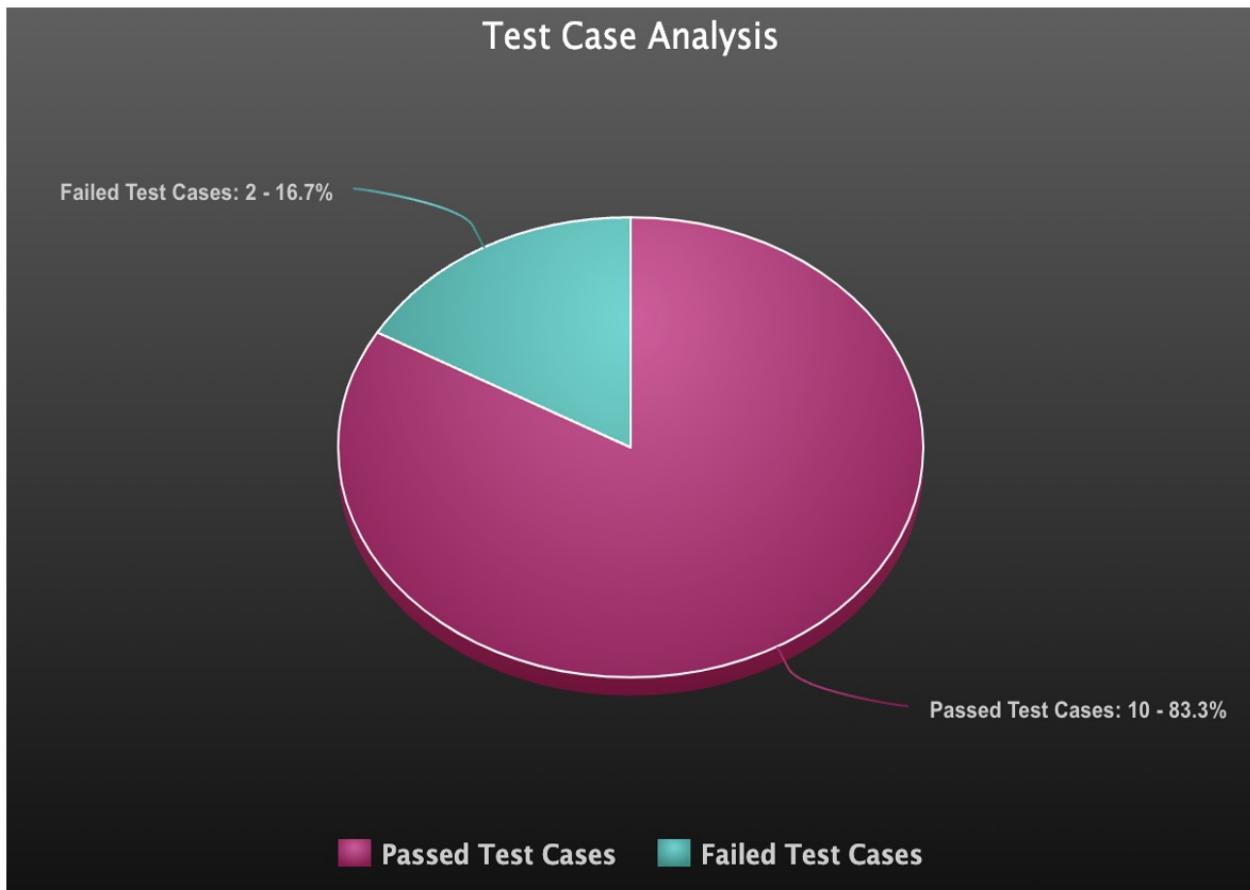
### Conventional Report

---

- Input Classification: 7
- Input Context: 6
- Total possible cases: 42
- No of used test cases: 12
- Test cases passed: 10
- Total cases failed: 2

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
Question Type	12	10	83.33%	2	16.66%
Sentiment					
<b>Total</b>	<b>12</b>	<b>10</b>	<b>83.33%</b>	<b>2</b>	<b>16.66%</b>

# Q/A Interaction: Conventional Test Analysis



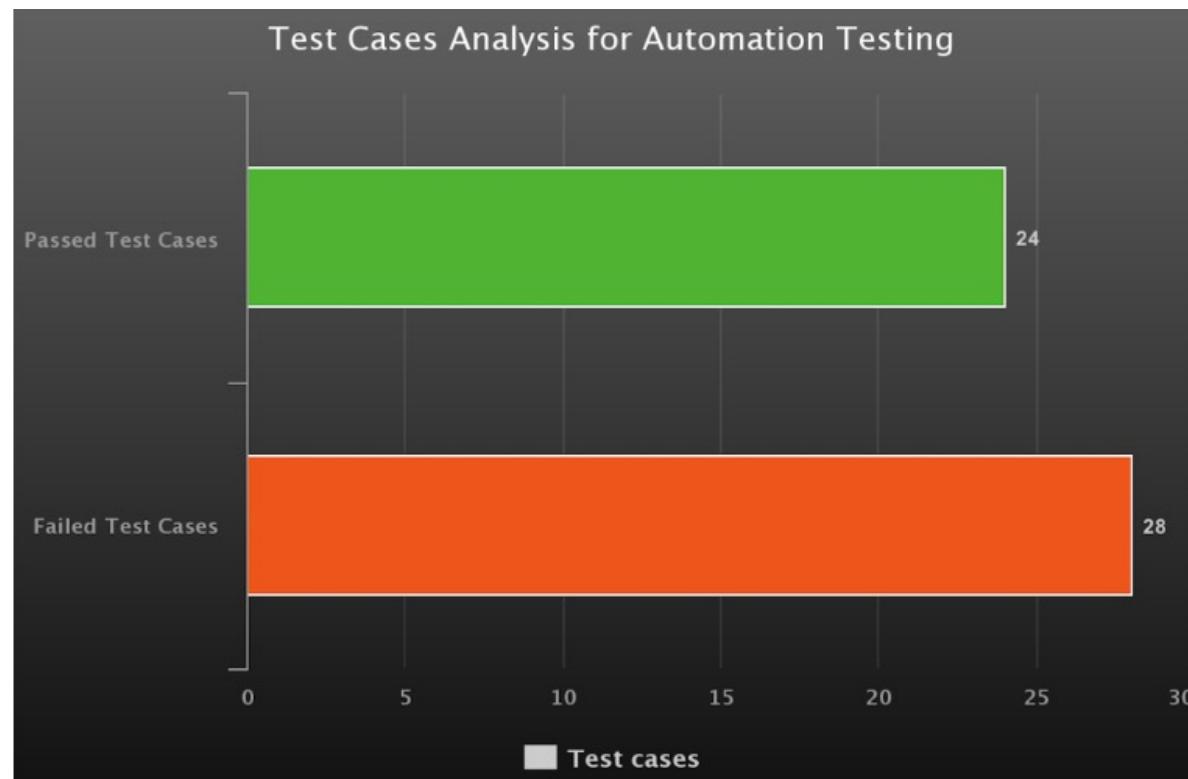
# Q/A Interaction: Automation testing

```
Test case 24:  
Input: Who invented @ sign  
Output: actually it was discovered by several people at about the same time  
False  
  
Test case 25:  
Input: Why you dont die 2 days from now  
Output: not too long perhaps due to noise in my circuits no i cannot die as i am made from software dying is something humans do not robots  
True  
  
Test case 26:  
Input: Why you so good are you?  
Output: its just how i am  
True
```

```
Test case 51:  
Input: How many hours of sleep is recommended for adults  
Output: most humans require 8 hours of sleep each night machines are better and never need to rest  
True
```

```
Total positive results: 24  
Total negative results: 28  
Total time consuming: 415
```

# Q/A Interaction: Automation Test Analysis



# Q/A Interaction:

## Conventional vs Automation testing

Q&A Interaction	Conventional	Automation
Total Test Cases	12	52
Passed Test Cases	10	24
Failed Test Cases	2	28
Time Consumed	15.7 minutes	6.9 minutes

# AI test complexity

- Domain Knowledge: 216 cases
- Chatbot Memory: 108 cases
- Chatbot Pattern: 144 cases
- Q/A Interaction: 108 cases
- Total: 576 cases



Thank You

