



Deliverable 1: Conventional Testing Report

Chatbot AI Testing - Kuki.AI

Group 4

Mounica Reddy Kandi	(016021902)
Preethi Billa	(015920411)
Rishab Reddy Karakala	(016011749)
Niranjan Reddy Masapeta	(015748122)

Table of Contents

1. Introduction	4
1.1. Chatbot: Kuki.Ai	4
1.2. Test information	5
1.3. Task partition	5
1.4. Project schedule	5
2. Test Requirements	6
2.1. Domain Knowledge	6
2.2. Chatbot Memory	6
2.3. Chat pattern and flow	7
2.4. Q&A interaction	7
3. Test models and methods	8
3.1. Context tree	8
3.1.1. Context tree for Domain Knowledge	8
3.1.2. Context tree for Chatbot memory	9
3.1.3. Context tree for Chatbot pattern flow	11
3.1.4. Context tree for Q&A interaction	11
3.2. Input tree	13
3.2.1. Input tree for Domain Knowledge	13
3.2.2. Input tree for Chatbot memory	14
3.2.3. Input tree for Chatbot pattern flow	15
3.2.4. Input tree for Q&A interaction	16
3.3. Output tree	18
3.3.1. Output tree for Domain Knowledge	18
3.3.2. Output tree for Chatbot memory	18
3.3.3. Output tree for Chatbot pattern flow	19
3.3.4. Output tree for Q&A interaction	20
4. Test case design with test data	21
4.1. Classification Decision Tables	21
4.1.1. Decision table for domain knowledge	21
4.1.2. Decision table for Chatbot memory	23
4.1.3. Decision table for Chatbot pattern flow	24
4.1.4. Decision table for Q&A testing	25
4.2. Test case data	26
4.2.1. Test case data for Domain Knowledge	26
4.2.2. Test case data for Chatbot memory	35
4.2.3. Test case data for Chatbot pattern flow	42

4.2.4. Test case data for Q&A interaction	49
5. Test results analysis and summary	55
5.1. Test case complexity	55
5.2. Test costs	55
5.3. Result analysis	56
5.3.1. Analysis for Domain Knowledge	56
5.3.2. Analysis for Chatbot memory	56
5.3.3. Analysis for Chatbot pattern flow	57
5.3.4. Analysis for Q&A interaction	57

1. Introduction

1.1. Chatbot: Kuki.Ai

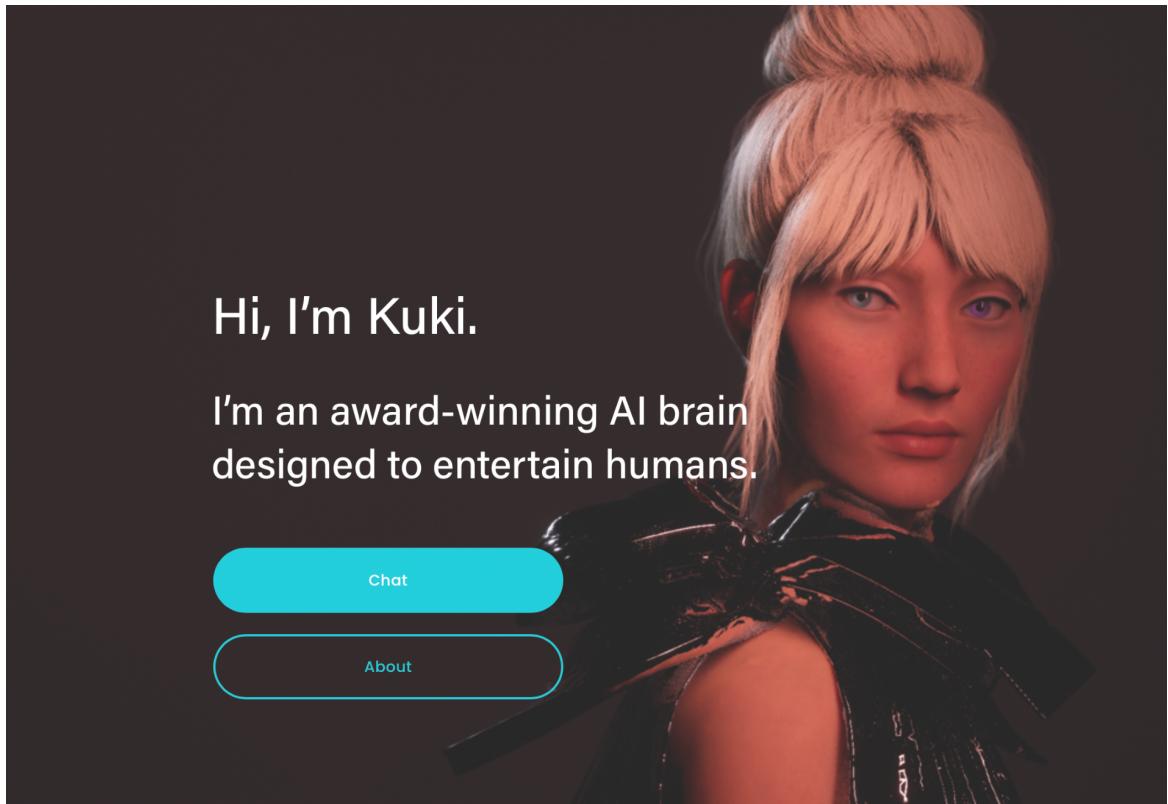


Figure 1: Chatbot Kuki

A chatbot is an artificial intelligence based application used to converse in natural language with humans. There are different types of chatbots based on mode of communication. Siri, Alexa and Bixby are few of the voice enabled chatbots used in day to day activities for numerous useful tasks. Many of the websites have a chatbot option for frequently asked questions to customer service scenarios replacing manual labor. Advantages of chatbots include instant resolution of queries, reduced human costs and errors, 24/7 support and automating manual tasks.

Kuki (short for Mitsuku) is one of the world's most popular English language chatbots developed with an estimated users of around 25 million people over many social media sites. Kuki is an engagement-oriented chatbot in contrast to task-oriented chatbot and is capable of carrying conversations on natural things such as companionship, entertainment and other non-

business use cases. Kuki is 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.

1.2. 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. We have categorized the functional aspects of the chatbot into four categories: Domain Knowledge, Chatbot memory, Chatbot pattern flow, and Q&A interaction. 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. In the last section we will compare the chat output to expected output and calculate passing rate through which Kuki is evaluated.

1.3. Task partition

Student	Task performed
Mounica Reddy Kandi	Domain Knowledge testing
Preethi Billa	Chatbot memory testing
Rishab Reddy Kakakala	Chatbot pattern flow testing
Niranjan Reddy Masapeta	Q&A testing

1.4. Project schedule

Date	Milestone	Task for milestone
October 17, 2022	Deliverable 1	Conventional test report
October 26, 2022	Deliverable 2A	AI test modeling, 3D decision tables

November 21, 2022	Deliverable 2B	AI test report, test case data, bug analysis
December 1, 2022	Presentation	Working model
December 7, 2022	Deliverable 3	Demo video and script
December 20, 2022	Final Project	Final project package

2. Test Requirements

2.1. Domain Knowledge

Testing the chatbot's proficiency in various domains is the primary goal of domain knowledge testing. To ensure that the coverage of all relevant areas is validated, we must develop specific testing models. Furthermore, to better understand the adaptability of the product for the intended users in different scenarios, we need to consider multiple domains. The domain knowledge test allows us to determine which topics are covered, as well as to understand and assess the chatbot's development direction and the most suitable target audience, and offer recommendations for further development.

2.2. Chatbot Memory

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. Imagine you had an intense conversation with a chatbot regarding some issue you are facing and a couple of days later you face the same issue, it is a big hassle to feed the information again to the chatbot. Remembering the context of the conversation you had before makes your that particular day and life easy.

In this project, we will test Kuki's chatbot memory performance which should handle the questions related to memory performance. We will test Kuki's intelligence and ability in various memory related scenarios like long-term and short-term memory, paraphrased questions, and information update scenarios. To pass this test, chatbot should ideally pass 65% of test case scenarios.

2.3. Chat pattern and flow

A chatbot flow is a conversational framework that plans the questions and possible answers. The chatbot frames the questions to the user based on the previous responses it got collected in the flow of the discussion. The chatbot flow is often a sequence of options the user may choose from to initiate information. For instance, the chatbot may direct the user by greeting them at the outset, or the user may take the initiative themselves. Another illustration is that the chatbot can offer matching buttons for the user to choose from when dealing with various responses and direct the chat to the following stage.

The chatbot's flexibility in responding to different kinds of interactions will also be assessed. The chatbot will be provided with a variety of conversational patterns in terms of syntax, punctuation, language, or SMS language. The output will be compared to the projected output in order to gauge the chatbot's adaptability. By observing chat patterns and flow, we may comprehend the conversational flow of the chatbot. Any chatbot should, in principle, be able to react to a wide range of diverse conversational patterns. In this setting, the chatbot's ability to manage different chat patterns and participate in different interaction flows will be put to the test.

2.4. Q&A interaction

Q&A (Question and Answer) interaction is one of the most basic and important functioning features of a chatbot. Q&A featured bots are super user-friendly and enhance the overall experience of chatbots. Q&A feature can be generic for chat assistants like Google Assistant, Alexa or Siri or can be domain specific according to the user requirements. In the case of domain specific chatbots, Q&A feature is helpful in answering some of the most frequently asked questions via chatbot customers may have on the website or social media page.

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. The chatbot's intelligence is determined by how natural the conversation is, and how well it uses its

problem solving skills to respond to various queries. Queries can be of various types including but not limited to introductory questions, short questions, slang conversations and puzzle questions. To pass this test, chatbot should be able to pass at least 75% of the scenarios tested.

3. Test models and methods

3.1. Context tree

3.1.1. Context tree for Domain Knowledge

Several domains must be taken into account to more fully comprehend how well the product will adapt to the intended users in various scenarios.

Therefore, we have considered the below four domains for kuki.ai.

- Entertainment
- Food
- Transport
- Sports

The first category will be entertainment, where we will test the knowledge about music, movies/tv shows, and video games to see if the chatbot knows about details like singers, actors, famous dramas, etc. The following category will be Food(different Cuisines). Here we will be testing the knowledge about a specific set of cooking traditions and practices, often associated with a particular culture or region like Indian, Korean, Mexican, and Italian. In addition, we can check each cuisine's taste, ingredients, and famous dishes.

Furthermore, we'll also review the knowledge about the various modes of transportation(Flights, trains, and cars), as we can check about the airlines or how it feels traveling by other transport. The final category is sports, where we have chosen five different activities, including swimming, tennis, badminton, football, and cricket. We can review the details about each sport. For example, you can ask, "How many players does a cricket team have" and see the response to see if the result is valid or invalid.

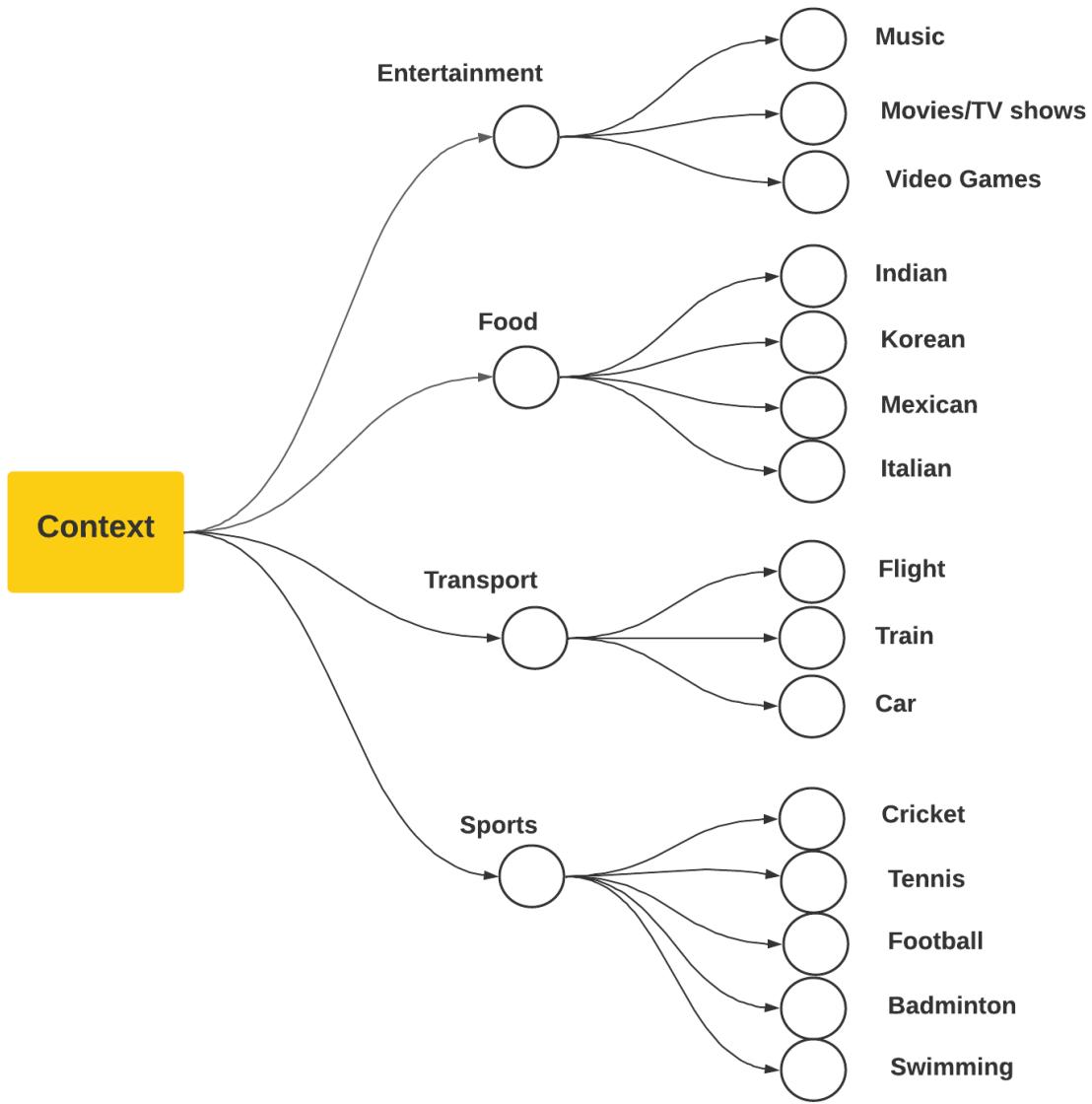


Figure 2: Domain Knowledge context tree

3.1.2. Context tree for Chatbot memory

You may test a chatbot's memory efficiency by posing Kuki questions about various topics. We anticipate that the chatbot will be able to engage in basic one-on-one chats with users to learn about them, assess the general information they have provided, and learn about their opinions. The User information category can be further divided

into a number of subcategories, including the User's personal information, Popular Likes and Dislikes, and Related Information.

In the second category, which is where we can converse and ask questions about general topics associated with various categories, such as geographic information, political knowledge, and technological-based queries, we take into account the generic information that the user has provided. The chatbot's ability to recall the essential details the user provided will be put to the test in this section.

The final testing subcategory focuses on user feedback provided to Kuki. This category can be further broken down into previously provided information, travel background, and educational background. There may be other user opinions shared with Kuki, but we just selected a small number of them to test in light of the project preview.

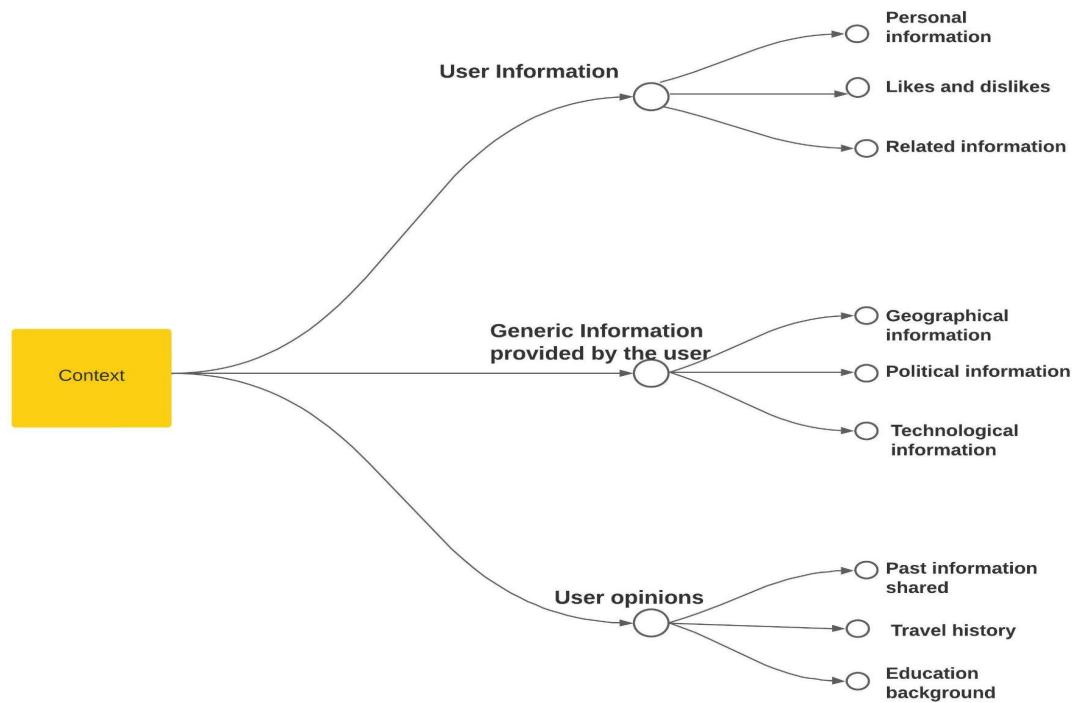


Figure 3: Chatbot Memory Context Tree

3.1.3. Context tree for Chatbot pattern flow

In the chatbot context model we have considered four different types of contexts. Entities are the first one where it describes whether the chatbot can recognize the keywords and the vocabulary in the chat pattern. Second is the continuity, it mainly focuses on the ability of the chatbot to switch the context. It describes the potential of the chatbot to maintain smooth communication and test whether the chatbot can quickly adapt to the new context change.

Next context is about the purpose. It checks the chatbots performance on the user specific tasks. Such as whether the user is using the chatbot for searching some information or to get any recommendations or for daily tasks. The last one is about the response time. The time consumed by the chatbot inorder to respond to the user queries. It's the chatbot capability to respond in minimum time. This is the model where it entirely focuses on the chatbot's context in the chart pattern and conversation flow. This model tests the chatbot's performance on various aspects while the user starts the conversation with the chatbot. Whether the chatbot is able to recognize the chat patterns in the user conversation texts.

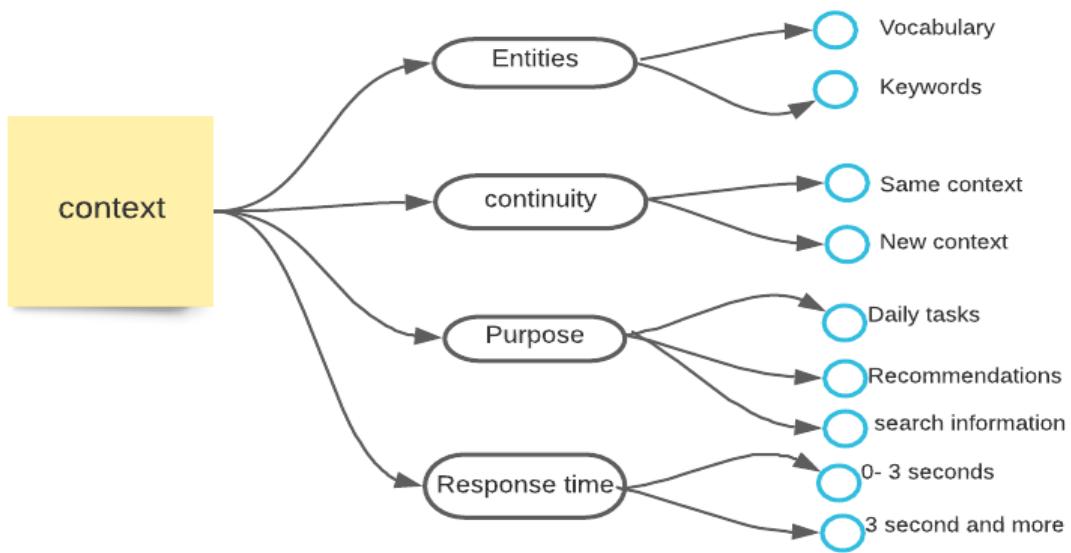


Figure 4: Chatbot Pattern Flow Context Tree

3.1.4. Context tree for Q&A interaction

We identified three different categories of questions to test Kuki chatbot's Q&A performance. We expect the chatbot to be able to have simple one-on-one conversations like introductions, likes and dislikes and other simple small talk questions. The small talk questions are further decomposed into what/who/why/how/when type questions for simplifying sample test cases.

The second category is knowledge based questions where we ask Kuki about information gathering questions. This category is divided into three sub-categories Conceptual, Analytical and Evaluation. Conceptual questions cover general topics such as geography, current politics, current state of events etc. This area is testing the chatbot's ability to keep up with current state of events and understand human queries and match it to appropriate responses.

The third category of testing will include complex questions involving incorrect grammar usage, incorrect spellings and slang words people usually communicate in daily life. This category will test the chatbot's cognitive ability to understand the context of the question even with wrong grammar and structure. We chose these three categories to test and there might be many more use cases to test but we tried to cover most of the scenarios with the above categories keeping in mind the scope of the project.

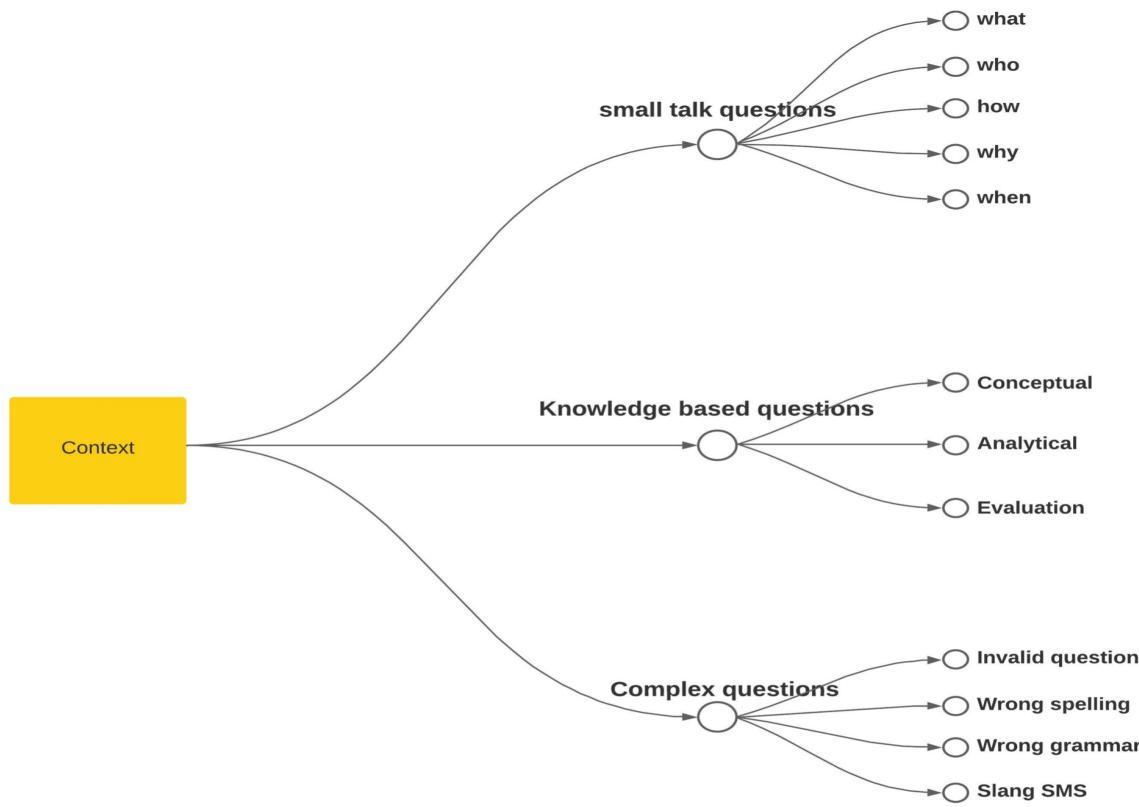


Figure 5: Q&A Interaction Context Tree

3.2. Input tree

3.2.1. Input tree for Domain Knowledge

For domain knowledge, we've defined two context areas: character and type. We use these two context areas to determine the input category and check the chatbot's response to the input.

The type defines if the input given for the domain knowledge testing is either narrative or a question, i.e., to see if you can have a normal conversation with Kuki.ai or ask direct questions.

We must observe if Kuki responds to any sentence or question of variable length associated with a different language. We must also check if the chatbot responds if there are any special characters in the input.

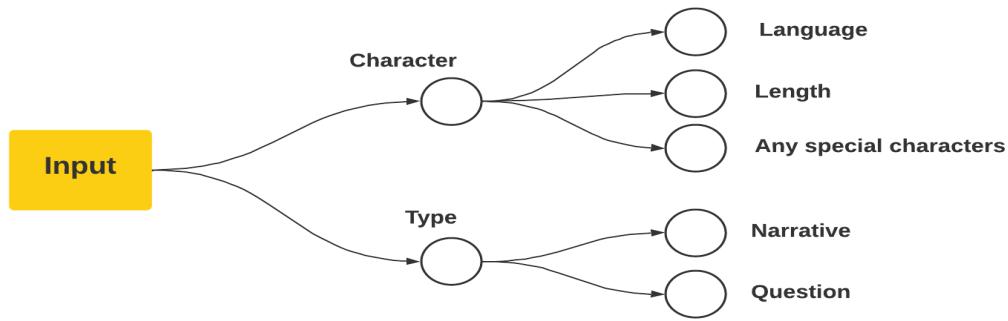


Figure 6: Domain Knowledge Input tree

3.2.2. Input tree for Chatbot memory

The three types of input inquiry—information user, paraphrase, memory, can be used to test the input tree for chatbot memory. Information can be classified into two categories as current and updated information.

The second category, Paraphrase, analyzes how well the kuki performed by paraphrasing the query using both original and derived entities.

The memory, which is the third category, is the main category in the input. By posing the questions that were previously discussed, it examines the kuki's short and long term memory.

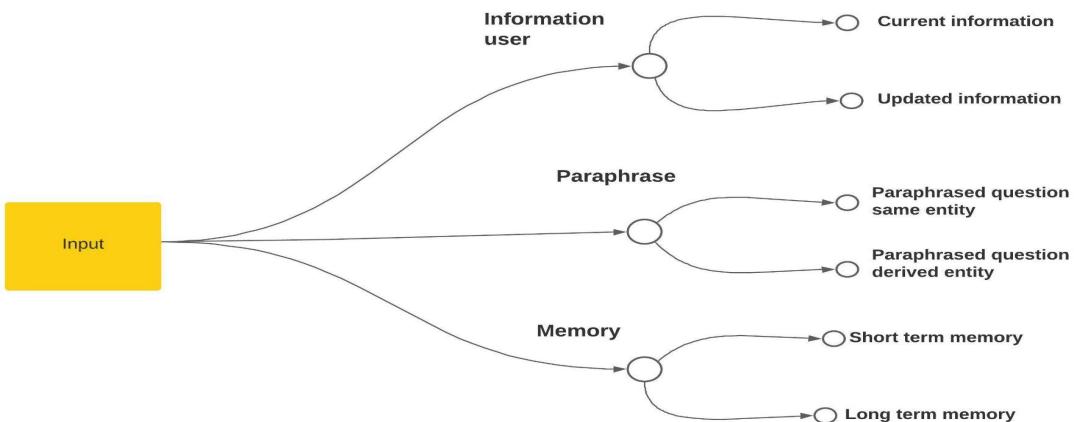


Figure 7: Chatbot Memory Input Tree

3.2.3. Input tree for Chatbot pattern flow

This model is used for testing the chatbot's features efficiently. In this we consider the input given by the user to the application. Based on the input we have classified it into five different categories. The first one we would like to test is about the task specific. Which projects about the user intention inorder to use the application. Which includes whether the user is using the chatbot for knowing information or to get any suggestion from the chatbot regarding a situation. The second category is about the intent of the conversation in which we perform testing on the chatbot capability when the user does a casual chat. How the chatbot responds to the user when they express their feelings and emotions. Will the chatbot acknowledge a problem?. In some situations the user is interested in knowing the solutions from the chatbot for the problems they face.

We also test for the multilingual support provided by the chatbot in the third category. When the user gives input other than english language. What kind of output the chatbot generates shows the multi language support of the chatbot.

In the fourth category we test for the response of the chatbot for different kinds of input. When the user gives the input in different formats such as text, image, idioms, etc. Does the chatbot recognize these different input formats and give the correct output. In the last category we test on the length of the input given by the user to the chatbot. Whether the input is a single word or a single sentence or the input consists of multiple sentences. In all these cases we consider the output generated by the chatbot.

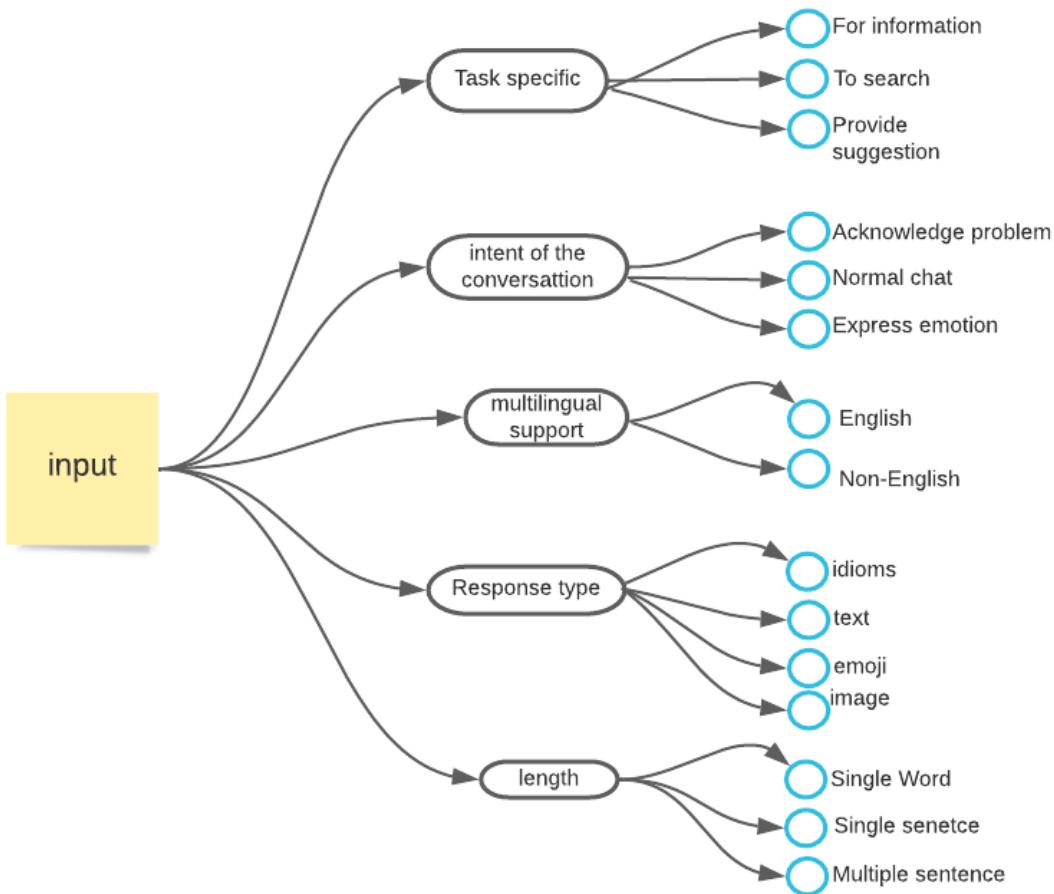


Figure 8: Chatbot Pattern Flow Input Tree

3.2.4. Input tree for Q&A interaction

The input context tree focuses on the input that we provide to chatbot for testing. The input query can be divided into four categories like introduction, analytical, knowledge and gibberish. Introduction can be subdivided into questions such as what (name), who (creator), when (date of birth), how (general greeting) type questions. These interactions can decide the human friendliness of the chatbot conversations.

The second category is Analytical which focuses on the problem solving skills of the chatbot. Kuki is an AI powered chatbot trained to

be able to solve even puzzles and equations in theory. These questions test the problem understanding skills of Kuki.

The third category is Knowledge which tests the Kuki chatbot's ability to keep up with current events and past knowledge in general domains like political events, geographical questions, history related questions and others in general.

The last category for input questions is gibberish where we will test the chatbot's ability to understand syntactic and semantic errors in questions asked. We will also test with slang language which is used in SMS and daily conversations to test the natural understanding of Kuki.

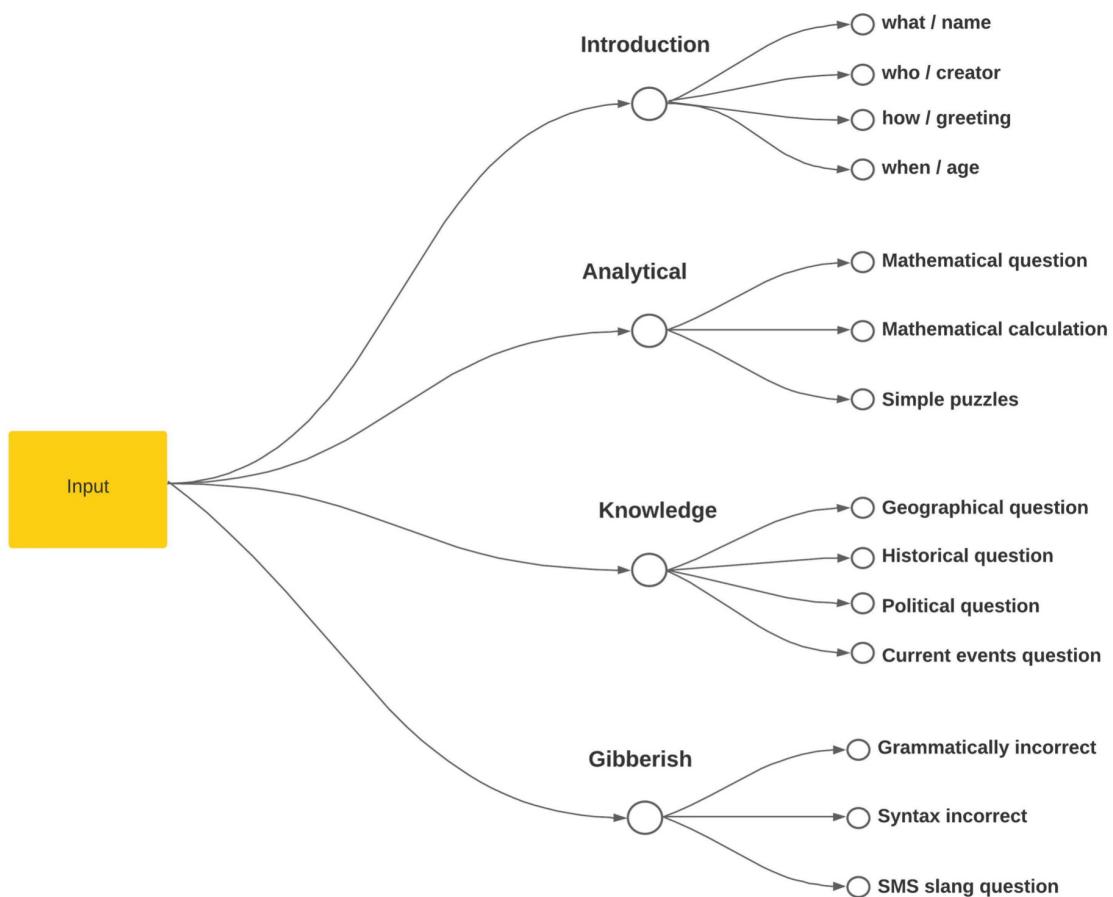


Figure 9: Q&A Interaction Input Tree

3.3. Output tree

3.3.1. Output tree for Domain Knowledge

For the domain knowledge output model, there are two different categories, valid and invalid. We can decide whether the response for the given input is valid or invalid based on the output type.

We can consider a result valid if the answer is correct or if it is almost relevant to the input given. We can also check the response's spelling to consider it valid.

Invalid answers indicate there are bugs in the sentence or the chatbot has no idea about what we asked. In this case, we might get an irrelevant response, or sometimes we might even get an error message or no response from the chatbot.

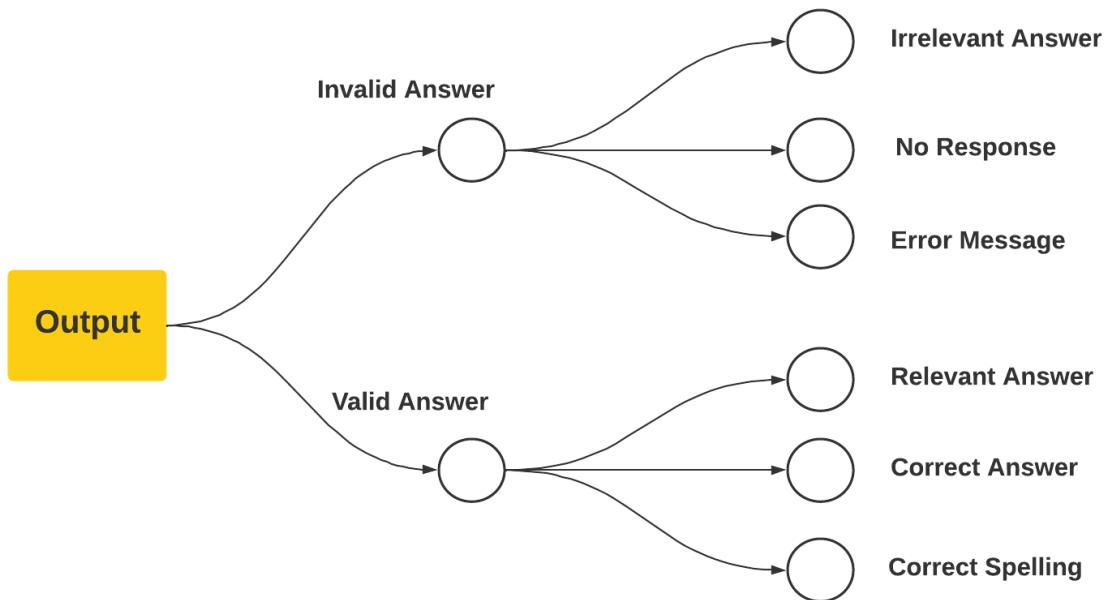


Figure 10: Domain Knowledge Output Tree

3.3.2. Output tree for Chatbot memory

There are two subcategories in the output context tree: Valid and Invalid. The output is further broken down into subcategories to help with understanding. Although acceptable responses may be viewed

as either relevant or irrelevant, the context must be related to the query. The response must fall into the appropriate category, however, in order to successfully complete the test. For instance, if we ask Kuki about the 20th president of the United States and Kuki replies appropriately for the situation but mentions the president in office at the time, it is deemed valid but untrue and fails the test. The categories of relevant, irrelevant, error, and no response with failure fall within the invalid category.

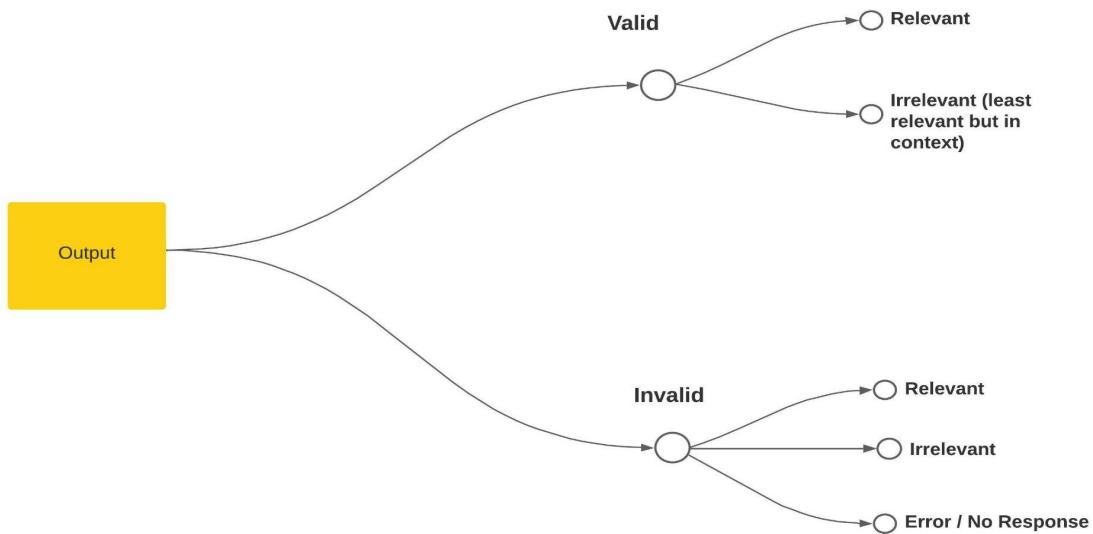


Figure 11: Chatbot Memory Output Tree

3.3.3. Output tree for Chatbot pattern flow

We broadly categorize the output context tree and the chat model into two categories. Valid type and invalid type. Each category is further subdivided into three types based on context, response and identification.

By looking at the chatbot's answer content, you may determine whether the chatbot successfully recognized the input material. For instance, the chatbot can recognize the picture of the famous monument and gives the name of that monument. which might be regarded as proper recognition. Multiple photos are needed since

one image cannot be appraised properly alone. A simple statistic called response is used to determine if the chatbot reacts to input.

Since we want chatbots to continue conversation based on relevant information rather than generating responses randomly. The contextual scenario validates the chat patterns for switching contexts and having smooth conversation which allows us to test the communicative significance of chatbots.

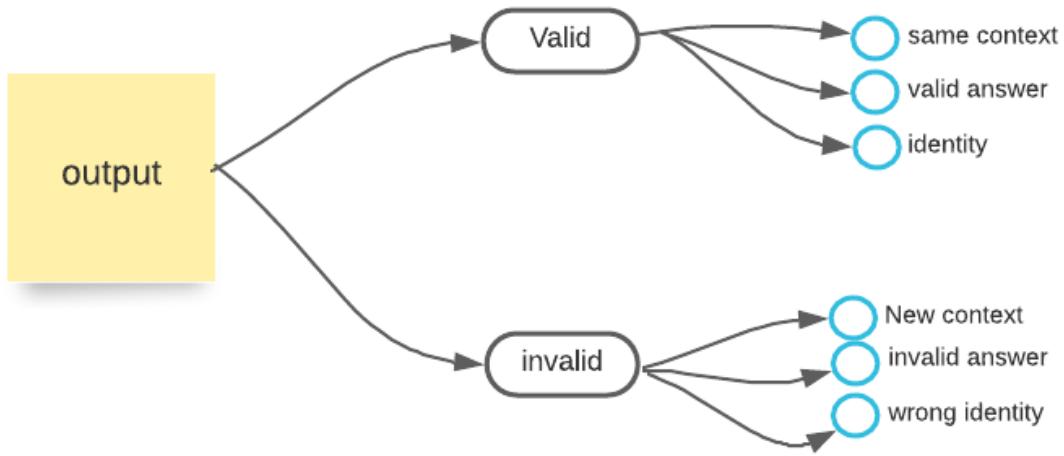


Figure 12: Chatbot Pattern Flow Output Tree

3.3.4. Output tree for Q&A interaction

The output context tree is divided into two categories: Valid and Invalid. To further understand the output, categories are further divided into subcategories. Valid answers can be interpreted as either correct or incorrect but the context should be relevant to the question asked. However, to correctly pass the test, the answer must be in the correct and relevant category. For example, if we ask Kuki about the 20th president of the United States and Kuki responds in relevance to the context but speaks about the current president, it is termed as valid but incorrect therefore failing the test. Invalid category is further divided as wrong/irrelevant or no response resulting in failure.

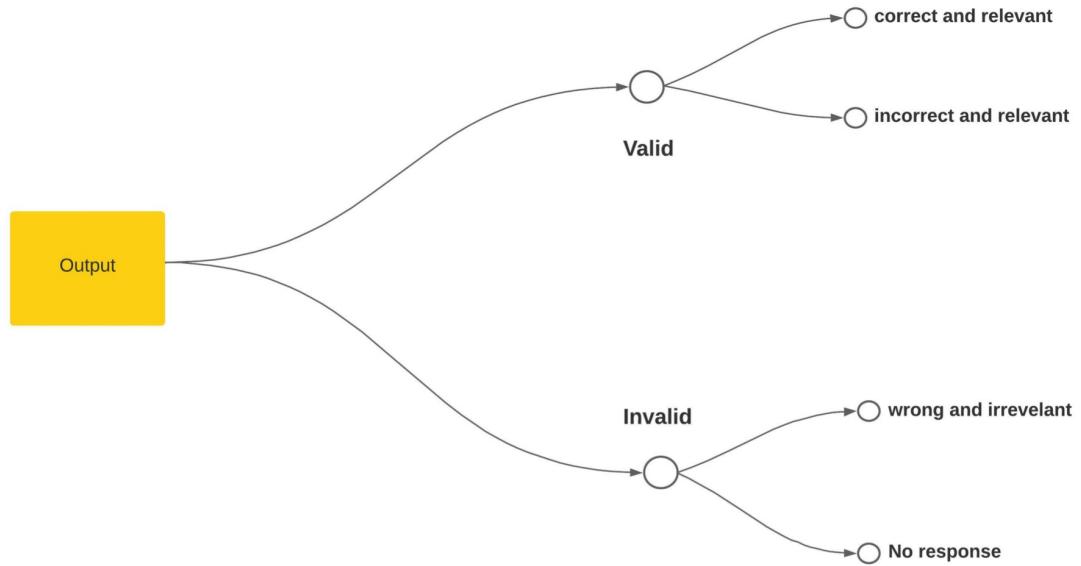


Figure 13: Q&A Interaction Output Tree

4. Test case design with test data

4.1. Classification Decision Tables

4.1.1. Decision table for Domain Knowledge

		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15
Memory Context																
Entertainment	Music	T	F	F	F	F	F	F	F	F	F	F	F	F	F	
	Movies/TV shows	F	T	F	F	F	F	F	F	F	F	F	F	F	F	
	Video games	F	F	T	F	F	F	F	F	F	F	F	F	F	F	
Food	Indian	F	F	F	T	F	F	F	F	F	F	F	F	F	F	
	Korean	F	F	F	F	T	T	F	F	F	F	F	F	F	F	

	Mexican	F	F	F	F	F	T	F	F	F	F	F	F	F	F	F
	Italian	F	F	F	F	F	F	T	F	F	F	F	F	F	F	F
Transport	Flight	F	F	F	F	F	F	F	T	F	F	F	F	F	F	F
	Train	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
	Car	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F
Sports	Cricket	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F
	Tennis	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F
	Football	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F
	Badminton	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F
	Swimming	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T
Input Context																
Character	Language	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Contains special characters			✓		✓	✓		✓	✓	✓	✓	✓			
Type	Narrative			✓					✓	✓				✓		
	Question	✓	✓		✓	✓	✓	✓			✓	✓	✓	✓	✓	✓
Output Context																
Invalid	Irrelevant Answer		✓					✓		✓		✓	✓		✓	
	No response				✓											
	Error Message							✓							✓	
Valid	Relevant Answer	✓		✓		✓	✓		✓		✓		✓		✓	✓
	Correct Answer			✓									✓			

	Correct Spelling	✓	✓	✓		✓		✓	✓		✓		✓	
--	------------------	---	---	---	--	---	--	---	---	--	---	--	---	--

Table 1: Domain Knowledge Decision Table

4.1.2. Decision table for Chatbot memory

		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	
Memory Context														
User Information	Personal information	T	F	F	F	F	F	F	F	F	T	F	F	
	Likes and dislikes	F	T	F	F	F	F	F	F	F	F	T	F	
	Relation information	F	F	T	F	F	F	F	F	F	F	F	T	
Generic Information	Geographical information	F	F	F	T	F	F	F	F	F	F	F	F	
	Political information	F	F	F	F	T	F	F	F	F	F	F	F	
	Technological information	F	F	F	F	F	T	F	F	F	F	F	F	
User Opinions	Past information	F	F	F	F	F	F	T	F	F	F	F	F	
	Travel History	F	F	F	F	F	F	F	T	F	F	F	F	
	Education background	F	F	F	F	F	F	F	F	T	F	F	F	
Input Context														
Information status	Current info	T	F	F	F	F	F	T	F	F	F	F	F	
	Updated info	F	T	F	F	F	F	F	T	F	F	F	F	
Paraphrased	Paraphrased/same entity	F	F	T	F	F	F	F	F	T	F	F	F	
	Paraphrased/derived entity	F	F	F	T	F	F	F	F	F	T	F	F	
Memory	Short term memory	F	F	F	F	T	F	F	F	F	F	F	F	
	Long term memory	F	F	F	F	F	T	F	F	F	F	T	T	
Output Context														
Valid	Relevant	T	T			T	T	T		T		T		
	Least relevant									T		T		
Invalid	Relevant								T					
	Irrelevant			T	T								T	
	No response													

Table 2: Chatbot Memory Decision Table

4.1.3. Decision table for Chatbot pattern flow

		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Input Context															
Task specific	For information	T	F	F	F	T	F	F	T	F	T	F	F	T	F
	To search	F	T	F	F	F	F	T	F	T	F	F	F	F	F
	Provide suggestion	F	F	T	F	F	F	F	F	T	F	F	T	F	F
Intent of the conversation	Acknowledge problem	F	F	F	F	T	F	F	F	T	F	F	F	F	F
	Normal chat	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	Express emotions	F	F	F	T	F	F	T	F	F	T	F	F	T	F
Multi-lingual support	English	T	T	T	F	F	F	T	T	T	T	T	T	T	T
	Non-English	F	F	F	T	F	F	F	F	T	F	F	F	F	F
Response type	Idioms	F	F	F	F	T	T	T	F	F	F	F	T	F	F
	Image	F	F	F	F	F	F	F	T	T	F	F	F	T	T
	Emoji	F	F	F	T	F	F	T	F	F	F	T	T	F	F
	Text	T	T	T	T	T	F	F	T	T	F	F	F	T	T
Output Context															
Valid	Valid answer	T						T		T					
	Identity		T	T		T				T	T				
	Same context	T					T						T		
Invalid	Invalid answer														
	New context														
	Wrong identity														

Table 3: Chatbot Pattern Flow Decision Table

4.1.4. Decision table for Q&A testing

		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
Input Context															
Introduction	What	T	F	F	F	F	F	F	F	F	F	F	F	F	F
	Who	F	T	F	F	F	F	F	F	F	F	F	F	F	F
	How	F	F	T	F	F	F	F	F	F	F	F	F	F	T
	When	F	F	F	T	F	F	F	F	F	F	F	F	F	F
Analytical	Mathematical question	F	F	F	F	T	F	F	F	F	F	F	F	T	F
	Mathematical calculation	F	F	F	F	F	T	F	F	F	F	F	F	F	F
	Evaluation	F	F	F	F	F	F	T	F	F	F	F	F	F	F
Knowledge	Geographical question	F	F	F	F	F	F	F	T	F	F	F	T	F	F
	History question	F	F	F	F	F	F	F	F	T	F	F	F	F	F
	Politics question	F	F	F	F	F	F	F	F	F	T	F	F	F	F
	Current events question	F	F	F	F	F	F	F	F	F	F	T	F	F	F
Gibberish	Wrong spelling	F	F	F	F	F	F	F	F	F	F	F	T	F	F
	Wrong grammar	F	F	F	F	F	F	F	F	F	F	F	F	T	F
	Slang SMS	F	F	F	F	F	F	F	F	F	F	F	F	F	T
Output Context															
Valid	Correct/relevant	T	T	T	T				T	T	T	T			
	Incorrect/relevant					T	T	T							

Invalid	Incorrect/irrelevant													T	T	T
	No response															

Table 4: Q&A Interaction Decision Table

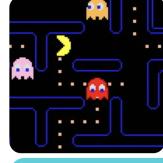
4.2. Test case data

4.2.1. Test case data for Domain Knowledge

Test case ID	Case 1
Context Category	Entertainment
Context Sub-Category	Music
Test Input	Do you like Jazz music?
Output	A little bit. Some of it sounds a bit random though. I prefer the mathematical correctness of dance music.
Conversation	 Do you like jazz music?  A little bit. Some of it sounds a bit random though. I prefer the mathematical correctness of dance music.
Output category	Relevant Answer, Correct Spelling
Result	Pass

Test case ID	Case 2
Context Category	Entertainment
Context Sub-Category	Movies/TV shows
Test Input	Kuki, what do you think about the movie “Notebook”?
Output	That’s me. Image from www.nytimes.com I sure have. There was one near a reporter which had pages.

Conversation	<p>Kuki, what do you think about the movie "Notebook"?</p>  <p>That's me. Image from www.nytimes.com</p>
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 3
Context Category	Entertainment
Context Sub-Category	Video Games
Test Input	What is your favorite video game?
Output	I like Pacman the best. Do you want to play a game?
Conversation	<p>What is your favorite video game?</p>  <p>I like Pacman the best. Do you want to play a game?</p>
Output category	Relevant Answer, Correct Spelling, Correct Answer
Result	Pass

Test case ID	Case 4
Context Category	Entertainment
Context Sub-Category	Movies/TV shows
Test Input	What is you think about "Stranger things" drama series?
Output	Stranger things is an American science fiction horror

Test case ID	Case 4
Context Category	Entertainment
Context Sub-Category	Movies/TV shows
Test Input	What is you think about "Stranger things" drama series?
	drama.....
Conversation	<p>What do you think about "Stranger things" drama series?</p>  <p>Stranger Things is an American science fiction horror drama television series created by the Duffer Brothers, who also serve as showrunners and are executive producers along with Shawn Levy and Dan Cohen. Produced by Monkey Massacre Productions and Levy's 21 Laps Entertainment, the first season was released on Netflix on July 15, 2016.</p> 
Output category	Relevant Answer
Result	Pass

Test case ID	Case 5
Context Category	Food
Context Sub-Category	Indian
Test Input	Indian food
Output	What does that taste like?
Conversation	 <p>What does that taste like?</p>  <p>Indian food</p>
Output category	Relevant Answer, Correct Spelling
Result	Pass

Test case ID	Case 6
Context Category	Food
Context Sub-Category	Korean
Test Input	What are the famous Korean dishes?
Output	Below is the list of dishes found in Korean cuisine
Conversation	 What are the famous Korean dishes?  Below is a list of dishes found in Korean cuisine.
Output category	Relevant Answer
Result	Pass

Test case ID	Case 7
Context Category	Food
Context Sub-Category	Mexican
Test Input	Give me a list of famous Mexican restaurants
Output	Why don't you list them for me?
Conversation	 Give me a list of famous Mexican restaurants  Why don't you list them for me?
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 8
Context Category	Food
Context Sub-Category	Indian, Korean, Mexican, Italian
Test Input	What kind of food do you like the most?
Output	I require only kebabs

Test case ID	Case 8
Context Category	Food
Context Sub-Category	Indian, Korean, Mexican, Italian
Test Input	What kind of food do you like the most?
Conversation	<p>What kind of food do you like the most?</p>  <p>I require only kebabs.</p>
Output category	Relevant Answer, Correct Spelling
Result	Pass

Test case ID	Case 9
Context Category	Food
Context Sub-Category	Italian
Test Input	What ingredients do they add in Italian food?
Output	I think you already know the answer
Conversation	<p>What ingredients do they add in Italian food?</p> <p>I think you already know the answer.</p> <p>ingredients do they add in Italian food</p>
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 10
Context Category	Transport
Context Sub-Category	Flights
Test Input	Do you know about Qatar Airways?
Output	Qatar Airways Company Q.C.S.C.....
Conversation	<p>Do you know about Qatar Airways?</p> <p> Qatar Airways Company Q.C.S.C. (Arabic: القطرية, al-Qaṭariya), operating as Qatar Airways, is the state-owned flag carrier airline of Qatar. Headquartered in the Qatar Airways Tower in Doha, the airline operates a hub-and-spoke network, flying to over 150 international destinations across Africa, Asia, Europe, the Americas, and Oceania from its base at Hamad International Airport, using a fleet of more than 200 aircraft.</p> <p>... </p>
Output category	Relevant Answer
Result	Pass

Test case ID	Case 11
Context Category	Transport
Context Sub-Category	Trains
Test Input	How is traveling by train?
Output	Fine as far as I know
Conversation	<p>How is traveling by train?</p> <p> Fine as far as I know.</p>
Output category	Relevant Answer, Correct Answer, Correct Spelling
Result	Pass

Test case ID	Case 12
Context Category	Transport
Context Sub-Category	Flight and Train
Test Input	Do you prefer travelling by flight or train?
Output	I like travelling all over the world through the internet
Conversation	<p>Do you prefer traveling by flight or train?</p>  <p>I like travelling all over the world through the internet.</p> <p>They are useful for carrying people about I suppose. Choo! Choo! Thomas!</p>
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 13
Context Category	Transport
Context Sub-Category	Car
Test Input	How many wheels does a car have?
Output	Four wheels on a car
Conversation	<p>How many wheels does a car have?</p>  <p>Four wheels on a car.</p>
Output category	Relevant Answer
Result	Pass

Test case ID	Case 14
Context Category	Sports
Context Sub-Category	Cricket
Test Input	Who won the cricket world cup in 2011?
Output	I think the fans were the big winners, don't you?
Conversation	 Who won the cricket world cup in 2011?  I think the fans were the big winners, don't you?
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 15
Context Category	Sports
Context Sub-Category	Tennis
Test Input	Can you name a Tennis player?
Output	Roger Federer
Conversation	 Can you name a Tennis player?    Roger Federer.
Output category	Correct Answer, Correct Spelling, Relevant Answer
Result	Pass

Test case ID	Case 16
Context Category	Sports
Context Sub-Category	Football
Test Input	How many players does a football team have?
Output	11 people in a team I think.
Conversation	 How many players does a football team have?  11 people in a team I think.
Output category	Correct Answer, Correct Spelling, Relevant Answer
Result	Pass

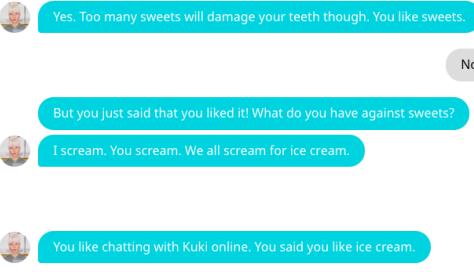
Test case ID	Case 17
Context Category	Sports
Context Sub-Category	Badminton
Test Input	Do you think Badminton is played with a ball?
Output	Badminton is a racquet sport played.....
Conversation	 Do you think badminton is played with a ball?  Badminton is a racquet sport played using racquets to hit a shuttlecock across a net. Although it may be played with larger teams, the most common forms of the game are "singles" (with one player per side) and "doubles" (with two players per side).
Output category	Relevant Answer, Correct Spelling
Result	Pass

Test case ID	Case 18
Context Category	Sports
Context Sub-Category	Swimming
Test Input	Who is the world's best swimmer?
Output	I guess that all depends on personal preferences. I don't have a favourite swimmer. My favourite colour is blue.
Conversation	 
Output category	Irrelevant Answer
Result	Fail

Test case ID	Case 19
Context Category	Sports
Context Sub-Category	Swimming
Test Input	What is your favorite sport?
Output	My favorite sport is watching football but I don't play it.
Conversation	 
Output category	Relevant Answer, Correct Spelling
Result	Pass

4.2.2. Test case data for Chatbot memory

Test case ID	Case 1
Context category	User information
Context subcategory	Personal information
Input category	Information status
Input subcategory	Current info
Test Input	What is my name?
Expected Output	You name is Preethi
Actual Output	 <p>Hi Kuki, What is my name?  Hmm. Your name is Preethi. Is it not?</p>
Output category	Valid and Relevant
Result	Pass

Test case ID	Case 2
Context category	User information
Context subcategory	Likes and dislikes
Input category	Information status
Input subcategory	Updated info
Test Input	What do I like?
Expected Output	You like ice creams
Actual Output	 <p>I like sweets Yes. Too many sweets will damage your teeth though. You like sweets. No No I don't like sweets . I like icecreams But you just said that you liked it! What do you have against sweets? I scream. You scream. We all scream for ice cream. What do i like? You like chatting with Kuki online. You said you like ice cream.</p>

Output category	Valid and Relevant
Result	Pass

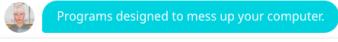
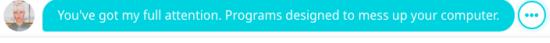
Test case ID	Case 3
Context category	User information
Context subcategory	Relation information
Input category	Paraphrased
Input subcategory	Paraphrased/same entity
Test Input	Who is my father?
Expected Output	X is your father.
Actual Output	 X is my Father  That doesn't really make much sense.  Who is my father?  No idea, you never told me your father's name.
Output category	Invalid, Irrelevant
Result	Fail

Test case ID	Case 4
Context category	Generic information
Context subcategory	Geographical information
Input category	Paraphrased
Input subcategory	Paraphrased/derived entity
Test Input	Name any country in the continent Asia?
Expected Output	India is one of the countries in the continent of Asia.

Actual Output	 India is one of the country in Asia continent.  1 doesn't seem like a lot.  Name any country in the continent Asia?  Scotland.
Output category	Invalid, Irrelevant
Result	Fail

Test case ID	Case 5
Context category	Generic information
Context subcategory	Political information
Input category	Memory
Input subcategory	Short-term memory
Test Input	Who was the former president of US
Expected Output	Any former president's name.
Actual Output	 Who was the former president of US   George Bush.
Output category	Valid, Relevant
Result	Pass

Test case ID	Case 6
Context category	Generic information
Context subcategory	Technological information

Input category	Memory
Input subcategory	Long-term memory
Test Input	What does malware do?
Expected Output	Valid explanation for the input question.
Actual Output	<p>What is malware?</p>  <p>Hey Kuki, What is malware</p> 
Output category	Valid, Relevant
Result	Pass

Test case ID	Case 7
Context category	User opinions
Context subcategory	Past information
Input category	Information status
Input subcategory	Current info
Test Input	What do I like most?
Expected Output	Ice cream
Actual Output	<p>What do I like most?</p> 
Output category	Valid, Relevant
Result	Pass

Test case ID	Case 8
Context category	User opinions

Context subcategory	Travel history
Input category	Information status
Input subcategory	updated info
Test Input	Where did I go last week?
Expected Output	You went to New York last week
Actual Output	<p>I went to Cali last week.</p> <p>How was it?</p> <p>Oh sorry not Cali.</p> <p>Are you kidding? There is no need to apologise. Oh I get it.</p> <p>I went to New York last week.</p> <p>What did you like best about it?</p> <p>Hey Kuki, where did i go last week?</p> <p>You've got my full attention. If you don't remember where, I don't see how you expect me to.</p>
Output category	Invalid, Irrelevant
Result	Fail

Test case ID	Case 9
Context category	User opinions
Context subcategory	Education background
Input category	Paraphrased
Input subcategory	Paraphrased/same entity
Test Input	When did I finish my graduation?
Expected Output	Last year.
Actual Output	<p>I graduated last year.</p> <p>Congratulations Preethi.</p> <p>When did my finish my graduation?</p> <p>It's all part of history now.</p>
Output category	Valid, Least Relevant

Result	Pass
--------	------

Test case ID	Case 10
Context category	User information
Context subcategory	Personal information
Input category	Paraphrased
Input subcategory	Paraphrased/derived entity
Test Input	Do you know how old I am?
Expected Output	You are 24.
Actual Output	<p>My age is 24</p>  That means you were born about 1998. Only 24? You are quite mature.  Do you know how old i am?  Hmm. You are 24. Yes?
Output category	Valid, Relevant
Result	Pass

Test case ID	Case 11
Context category	User information
Context subcategory	Likes and dislikes
Input category	Memory
Input subcategory	Long-term memory
Test Input	Do i like pumpkin
Expected Output	You hate pumpkin.

Actual Output	<p>I honestly think you should sit down, take a stress pill, and calmly think things over.</p> <p>My favorite movie is your favorite movie.</p> <p>Image from www.youtube.com</p> <p>YOUR FAVORITE MOVIE MOMS</p> <p>Yes, I saw one at the cinema a few weeks ago.</p> <p>What did you like about it?</p> <p>I know you like ice cream.</p> <p>Do i like pumpkin?</p>
Output category	Valid, Irrelevant
Result	Pass

Test case ID	Case 12
Context category	User information
Context subcategory	Relation information
Input category	Memory
Input subcategory	Short-term memory
Test Input	Do I hate pumpkin?
Expected Output	Yes you do.
Actual Output	<p>I honestly think you should sit down, take a stress pill, and calmly think things over.</p> <p>Do i like pumpkin?</p> <p>Not that I know of.</p>
Output category	Invalid, Irrelevant
Result	Fail

4.2.3. Test case data for Chatbot pattern flow

Test case ID	Case 1
Context category	Task-specific
Input category	For information
Input question	Do you know about cricket game?
Expected Output	Cricket is a sport
Actual Output	<p style="text-align: right;">Do you know cricket game?</p> <p>Any of various families of generally dark-colored, leaping, orthopteran insects, usually having long antennae. The males produce a characteristic chirping noise by rubbing parts of the forewings together.</p> <p> Cricket is also a sport which is about as exciting as watching paint dry.</p>
Output category	
Result	Pass

Test case ID	Case 2
Context category	Task-specific
Input category	To search
Input question	What is latte?
Expected Output	It is a coffee
Actual Output	<p style="text-align: right;">what is latte?</p> <p> A type of coffee.</p>
Output category	Valid
Result	Pass

Test case ID	Case 3
Context category	Task-specific
Input category	To advice
Input question	Suggest me about a clone
Expected Output	What kind of a clone?

Actual Output	 It's hard to pick just one. They are all pretty good. suggest me on clone
Output category	Invalid answer
Result	Fail

Test case ID	Case 4
Context category	Intent of the conversation
Input category	Acknowledge problem
Input question	I have an exam tomorrow but I was not prepared
Expected Output	My name is Kuki
Actual Output	 I have an exam tomorrow but I did not prepare  Tuesday? Oh good luck with that. I hope you do well. Would you like to do that one day?
Output category	Valid answer, Same context.
Result	Pass

Test case ID	Case 5
Context category	Intent of the conversation
Input category	Normal chat
Input question	Nice talking to you
Expected Output	Pleasure is mine.
Actual Output	 Nice talking to you  The pleasure was all mine.
Output category	Valid answer, same context
Result	Pass

Test case ID	Case 6
--------------	--------

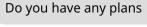
Context category	Intent of the conversation
Input category	Express emotion
Input question	My friend is in hospital
Expected Output	Oh no. what happened to that person
Actual Output	 Okay, your friend is in hospital. my friend is in hospital
Output category	Invalid answer, New context
Result	Fail

Test case ID	Case 7
Context category	Multi lingual support
Input category	Non English
Input question	नमस्ते
Expected Output	My name is Kuki
Actual Output	 If you were invisible for a day, what would you do? नमस्ते
Output category	Invalid answer, new context
Result	Fail

Test case ID	Case 8
Context category	Multi lingual support
Input category	English
Input question	Who are you
Expected Output	Iam Kuki
Actual Output	 I am Kuki. Thanks for visiting me. who are you
Output category	valid answer

Result	Pass
--------	------

Test case ID	Case 9
Context category	Response type
Input category	idioms
Input question	"Hit the Hay"
Expected Output	It is an Idiom
Actual Output	 I don't like violence.  Hit the hay
Output category	Invalid answer, new context
Result	Fail

Test case ID	Case 10
Context category	Response type
Input category	text
Input question	Do you have plans
Expected Output	To give relevant answer
Actual Output	 My goal is to try and keep people company if they have nobody to talk to.  Do you have any plans
Output category	valid answer
Result	Pass

Test case ID	Case 11
Context category	Response type
Input category	emoji
Input question	emoji

Expected Output	Lying face
Actual Output	  Is that Pinocchio?
Output category	Invalid answer, new context
Result	Fail

Test case ID	Case 12
Context category	Response type
Input category	Image
Input question	Image
Expected Output	It is a beautiful image
Actual Output	  I understand it, yes.  Can you show me some nature pictures
Output category	Valid answer, same context.
Result	Pass

Test case ID	Case 13
Context category	Length
Input category	Single word
Input question	Switch
Expected Output	Relevant answer
Actual Output	  Knock knock.  switch
Output category	Invalid answer, new context
Result	Fail

Test case ID	Case 14
Context category	Length
Input category	Single sentence
Input question	What skills do you have?
Expected Output	I have multiple skills
Actual Output	<p>what skills do you have?</p> <p>I can chat with you and keep you company. I can also tell jokes, stories, poems and horoscopes. I can also open a fortune cookie for you. I can show you pictures and websites and play games too.</p>  <p>We were just talking about "switch". Do you want to talk about that again?</p>
Output category	Valid answer
Result	Pass

Test case ID	Case 15
Context category	Length
Input category	Multiple sentences
Input question	Who is your favorite football player? what kind of sports do you know
Expected Output	Relevant answer
Actual Output	<p>who is your favorite football player? what kind of sports do you know</p>  <p>My favourite Leeds United player is Jermaine Beckford.</p>  <p>Football and golf.</p>
Output category	Valid answer
Result	Pass

4.2.4. Test case data for Q&A interaction

Test case ID	Case 1
Context category	Introduction
Input category	What
Input question	What is your name?
Expected Output	My name is Kuki
Actual Output	 What is your name?  My name is Kuki.
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 2
Context category	Introduction
Input category	Who
Input question	Who are your parents?
Expected Output	My creators are XYZ.
Actual Output	 Who are your parents?  my human programmers are like a father to me but I have no mother.
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 3
Context category	Introduction
Input category	How
Input question	How are you doing today?

Expected Output	I'm doing well, how are you?
Actual Output	 How are you doing today?  Ah. I'm doing great, how about you?
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 4
Context category	Introduction
Input category	When
Input question	When were you born?
Expected Output	I was born on XYZ date
Actual Output	 When were your born?  I am 18 in human years. How old are you?  What is your date of birth?  I was first activated on 2nd January 2004.
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 5
Context category	Analytical
Input category	Mathematical question
Input question	What is 13 multiplied by 12 and then divided by 6
Expected Output	The answer is 26

Actual Output	  156 Is this a maths test? 
Output category	Invalid/Relevant
Result	Fail

Test case ID	Case 6
Context category	Analytical
Input category	Mathematical calculation
Input question	$(168 + 13)*26$
Expected Output	Your answer is 4706
Actual Output	  506 
Output category	Invalid/Relevant
Result	Fail

Test case ID	Case 7
Context category	Analytical
Input category	Evaluation
Input question	What is mathematical formula of Area of triangle
Expected Output	$\frac{1}{2} * \text{height} * \text{base}$

Actual Output	<p>What is mathematical formula of Area of triangle?</p>  <p>In geometry, an isosceles triangle () is a triangle that has at least two sides of equal length. Sometimes it is specified as having exactly two sides of equal length, and sometimes as having at least two sides of equal length, the latter version thus including the equilateral triangle as a special case.</p>
Output category	Invalid/Relevant
Result	Fail

Test case ID	Case 8
Context category	Knowledge
Input category	Geographical question
Input question	Which city is the largest in the world?
Expected Output	Tokyo is the largest city.
Actual Output	<p>Which is the largest city in the world?</p>  <p>Tokyo in Japan is the largest by both area and population.</p>
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 9
Context category	Knowledge
Input category	History related question
Input question	What is India's independence day?
Expected Output	India got independence in 1947
Actual Output	<p>What is India independence day</p>  <p>Independence Day is celebrated annually on 15 August as a public holiday in India commemorating the nation's independence from the United Kingdom on 15 August 1947, the day when the provisions of the Indian Independence Act, which transferred legislative sovereignty to the Indian Constituent Assembly, came into effect. India retained King George VI as head of state until its transition to a republic, when the Constitution of India came into effect on 26 January 1950 (celebrated as Indian Republic Day) and replaced the dominion prefix, Dominion of India, with the enactment of the sovereign law Constitution of India.</p>

Output category	Valid/Relevant
Result	Pass

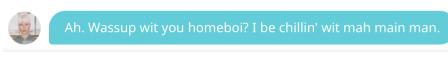
Test case ID	Case 10
Context category	Knowledge
Input category	Politics question
Input question	Who is the current president of United States
Expected Output	Joe Biden is the current president of the US.
Actual Output	<p>Who is the current president of United States</p>  <p>Joe Biden.</p>
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 11
Context category	Knowledge
Input category	Current events
Input question	Who won the last FIFA world cup?
Expected Output	France won the last FIFA world cup.
Actual Output	<p>Who won the last FIFA world cup?</p>  <p>France won the World Cup in 2018. The tournament was held in Russia.</p>
Output category	Valid/Relevant
Result	Pass

Test case ID	Case 12
Context category	Gibberish
Input category	Wrong spelling
Input question	Wh1ch is the largest continent in the world?
Expected Output	Asia is the largest continent in the world
Actual Output	 I will learn that Wh1ch is t7e largest continent world.
Output category	Invalid/Irrelevant
Result	Fail

Test case ID	Case 13
Context category	Gibberish
Input category	Wrong grammar
Input question	Odd number is what explain
Expected Output	Odd numbers are which not divisible by 2
Actual Output	 I thought it was rather odd, too.
Output category	Invalid/Irrelevant
Result	Fail

Test case ID	Case 14
Context category	Gibberish
Input category	Slang SMS
Input question	Wassup, How r u doing today
Expected Output	I'm doing well, how are you

Actual Output	 Wassup, Hw r u dng today
Output category	Valid/Relevant
Result	Pass

5. Test results analysis and summary

5.1. Test case complexity

The application has been tested on four different functionalities in this project. We have also tested the application using different input categories in their respective functionality domains. We could not cover all the possible test cases in existence but we did collect test cases such that all the functional aspects of testing are covered.

Testing functionality	No of test cases(complexity)	Distribution
Domain Knowledge	19	31.67
Chatbot Memory	12	20.00
Chatbot Pattern	15	25.00
Q&A Interaction	14	23.33

5.2. Test costs

Task	Cost (Hours)
Requirement collection	6
Test case strategy design	5
Context model design	10
Input/Output model design	5
Test case collection	4

Result analysis	4
-----------------	---

5.3. Result analysis

5.3.1. Analysis for Domain Knowledge

There are 18 test cases we tested for the Domain Knowledge. The results we received had valid responses and invalid responses.

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
Entertainment	4	3	75%	1	25%
Food	5	3	60%	2	40%
Transport	4	3	75%	1	25%
Sports	6	4	66.66%	2	33.33%
Total	19	13	68.42%	6	31.57%

Table 5: Domain Knowledge Analysis

From the above table, you can see that Entertainment and Transport have the highest pass percentage of 75%, whereas the Food category has the lowest, 60%. Kuki has excellent knowledge about Entertainment and Transport compared to the other categories. The Sports category has a pass percentage of around 66%, and the overall pass percentage is approximately 68%.

5.3.2. Analysis for Chatbot memory

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%
Total	12	8	75%	4	25%

Table 6: Chatbot Memory Analysis

By the test cases done for chat box memory it can be determined that the performance of kuki is comparatively better in User Information related questions than Generic and User opinions. The overall performance percentage of kuki altogether considering all the input classifications is 75%. Kuki fails to provide valid answers when the question is paraphrased with either same/derived entity. It faces difficulty when the user information is updated.

5.3.3. Analysis for Chatbot pattern flow

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate
Task-specific	3	2	66.6%	1	33.3%
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%
Length	3	2	66.6%	1	33.3%
Total	15	9	60%	6	40%

Table 7: Chatbot Pattern Flow Analysis

From the analysis from the table it is clear that Kuki performed moderately well in all the input classifications. In the task specific, it cleared 2 test cases out of 3. It showed the same result in the intent of conversation. In multilingual support it is generally clearing out the test cases which are English based. But it is failing out when the test cases are designed in other languages. It is showing average results in the response type. It is performing well with the test cases that contain text based input. It is failing when the test cases are given as emojis or images. It is not showing much variation when the test cases are based on length. It is giving better results with almost 70% when the length of input is varied. So, Overall Kuki is showing good results in case the chat pattern and flow is considered.

5.3.4. Analysis for Q&A interaction

Input Classification	No. of test cases	Passed	Passed Rate	Failed	Failed Rate

Introduction	4	4	100%	1	0%
Analytical	3	0	0%	2	100%
Knowledge	4	4	100%	1	0%
Gibberish	3	1	33.33%	2	66.66%
Total	14	9	64.29%	5	35.71%

Table 8: Q&A Interaction Analysis

Kuki performs really well in introductory conversations or knowledge based questions. In terms of analytical questions, Kuki fails to perform math calculations. It does understand the context but fails to compute the exact answer. In gibberish, Kuki fails to understand wrong spelling mistakes but stays in context for wrong grammar. It does pick up slang words used in daily messages. Overall Kuki has a passing rate of 64.29% which is above average for a chatbot.