Chatbot Testing: Challenges and approach for testing Chatbot applications

Author: Rajni Singh & Neeraj Jain
Agenda

**1. Context**
Understanding Chatbots and real time examples of uncertainties

**2. Challenges**
Challenges in testing chatbot

**3. Testing Approach**
Scope, Approach & Tools and Use case
01

Context

Understanding Chatbots, the evolution and its challenges
How it works?

User input in plain language:
- Text
- Speech

Channels allows user to connect to chatbot application

Chatbot is a “computer program” which
- Maintains a sensible and contextual conversation with a human
- Operates on cognitive services such as NLP is used to make it more intelligent.
- Processes business logic and prepare response to user input
- Can take actions
Types of Chatbot
Chabot can be broadly categorized in two categories.

AI-BASED
Built with artificial intelligence technique to understand the human language together with sentiments that does not rely on scripted conversation. It accepts free form of input and responds based on the existing domain knowledge with self-learning improvement with maturity.

RULE BASED
Chatbot behavior is based on a set of rules, flows, and triggers to respond to very specific commands being asked by the user. The conversation is usually scripted and chatbot responds to each question with a predefined rule and each step is picked with an explicit option. A simple example might be a chatbot that tells you the stock price on a given date.
Uncertainties: Why chatbot needs to be tested?

Facebook Inc.’s chatbots hit a 70% failure rate and zero emotional intelligence.

“Tay's responses have turned the bot into a joke, but they raise serious questions.”

In 2016, users were finding it impossible to unsubscribe WSJ

Ticketmaster Breach Traces to Embedded Chatbot Software
02 Challenges

What could go wrong in Chatbot testing?
Dependency on humungous amount of data

There is no defined input / output

Difficult to predict all scenarios

Continuous monitoring

Continuous learning from past behavior

AI applications are non-deterministic and probabilistic

Shift from traditional testing to AI testing
Chatbot testing challenges

- Uncertainty of user conversation
- Domain specific validations
- Multichannel user experience
- Chatbot intelligence not being questioned
- Chatbot security being compromised
- Lack of robust testing strategy
Approach

Scope, parameters, and testing tools
Chatbot testing scope

- On Boarding & Personality
- Context remembrance and switching
- Specialized testing
- Regression
- API
- Conversation flow
- Domain Specific
- Crowd
- Performance & Security
- NLP & Cognitive Services

Enhanced User Experience
Parameters for testing

- Security
- Personality
- Navigation
- Onboarding
- Error Management
- Intelligence
- Speed
- Accuracy
Testing tools

**BOX**

NLP model test and real time monitoring

**BOTIUM**

Conversational flow testing

**Test My Bot**

Automated Testing for Chatbots

It includes tools for recording and replaying conversations and integrates with CT/CI/CD

**ChatbotTest**

An open source guide that helps you identify chatbot's design issues under 7 different categories
Meet Ginger
Our in-house customized personal assistant
Ginger says:
- Notification about important event, conferences, announcement

Ginger bytes:
- Actionable insight that is customized to your role

Ask Ginger:
- Any question about Nagarro
Hello Rajni

I'm taking my first tender steps as your virtual assistant. Ask me a question about Nagarro and I'll try and help you with whatever I've learnt so far.
Beyond traditional testing for AI-infused Chatbots
End-to-end automation testing framework

Cloud/On-Premise Solution

Conversational Management
- Context
- Session
- User

Request
Response

Linguistic Framework
- Intent
- Entity
- Synonyms
- Action Binding

Model
NLP

Admin
- Configuration
- Security
- Training

Integration
Live Connectors
Enterprise Services

Training Set
Data import-export
Measure the performance of the bot

- Goal completion rate
- Self-service rate
- AI and ML rates
- User retention rate
- Fallback rate
NLP-based model life-cycle of the Chatbot

Performance Monitoring

Data Model Manager

Performance Tuning

BAF

RTM

Training

Performance

Monitoring

Chatbot Model

Client Questions

Client Using Chatbot

Re-Train

Measure

Measure

Read Client Interaction

Watson

LUIS

Dialog flow

RASA

others

© 2019 Nagarro – All rights reserved
Advanced security testing

1. End-to-end encryption
2. Two-factor authentication
3. User authentication
4. Intent authorization
5. Channel authentication
6. Compliance validation
7. Authentication timeout
8. Self-destructing messages
Domain-specific testing

This testing should start with identifying the domain of the chatbot and then identify the keywords of that domain.

A Chatbot built for IT should treat the word “Selenium” as an automation tool while a chatbot built for pharma industry should treat the “Selenium” as chemical.
Recommendations

- Monitor, measure and update (NLP, conversational etc.)
- Automate end to end (text, voice, etc.)
- Domain-specific testing
- Non-functional testing