Getting Chatbots Right

Assuring a Quality Customer Experience

Michael Monegan
VP, Product Management
April 10, 2018
SpeechTEK / Wash DC
Chatbots are coming! Chatbots are coming!

It’s still early days, but bots have strong appeal

• Instant gratification for consumers, with no puzzle-solving in self-service or waiting for an agent
• Conversational engagement that offers intuitive access to information and functionality
• Cool, emerging technology that can ultimately serve consumers better (not just faster)
• Builds on success of messaging platforms (Facebook, Amazon, Apple, WhatsApp, WeChat, Microsoft, etc.)

They’re expected to become unavoidably popular

• They fill gaps left uncovered by increasingly complex web apps and slow, inconsistent agents
• They drive self-service cost savings
• They can reduce task resolution time
• They feel natural
Chatbots are coming! Chatbots are coming!

The Problem: There are many new potential points for failure to avoid

The Real Problem: If we screw up The Chatbot Experience…

…the reaction may eclipse the worst days of consumer IVR avoidance:
What’s so hard about High Quality Bots?

1. They’re way more conversational. And each user brings an element of non-determinism to an interaction model that’s far less structured than DTMF and directed dialog speech. Examples:
   • A unique way of choosing words to express themselves
   • A context that influences the order the conversation takes, and that slots get filled
   • An environment and device that drives their selection of voice, touch or keyboard input
   • Questions, clarifications, corrections, new topics, etc.

2. Remembering and applying context is expected, but hard
   • Bots are being measured against the user’s friends
   • Personalization happens naturally

3. Bot deployments are more interwoven with web sites, Alexa, Messenger, chat agents, call centers, etc.

4. Performance is critical – managing responsiveness in UI is paramount to the users that prefer bots

5. Underlying cloud-based services are evolving fast – constant training and improvement leads to learning in the cloud (usually good, but not always)
Chatbots – What could possibly go wrong?

<table>
<thead>
<tr>
<th>Challenge Area</th>
<th>Example Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening Well</td>
<td>• Voice interface not transcribing accurately enough</td>
</tr>
<tr>
<td>Understanding</td>
<td>• Semantics mis-interpreted (wrong intent, missed entity extraction)</td>
</tr>
<tr>
<td></td>
<td>• Bad training data</td>
</tr>
<tr>
<td>Handling Prior Context</td>
<td>• Lack of faithful context transfer from web site or search launch point</td>
</tr>
<tr>
<td></td>
<td>• Didn’t use context the user expressed earlier or in a prior chat session</td>
</tr>
<tr>
<td></td>
<td>• Clinging to context that the user has mentally abandoned</td>
</tr>
<tr>
<td></td>
<td>• Too much learning – bot “personality” strongly influenced by other users</td>
</tr>
<tr>
<td></td>
<td>(e.g. MSFT Tay)</td>
</tr>
<tr>
<td>Escalation to Agent (Chat or Voice)</td>
<td>• Lack of faithful context transfer to a human chat agent</td>
</tr>
<tr>
<td></td>
<td>• Slow escalation to an available, appropriately skilled chat agent</td>
</tr>
<tr>
<td></td>
<td>Mishandling user engagement when no agents are available (i.e. manage expectations</td>
</tr>
<tr>
<td></td>
<td>with ETA)</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>• External services are slow and delay not managed well with user</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
How do we respond?

Our Top Ten Recommendations
Making Chatbots Great – A Top 10 List

Our recommendations for a 10-step Lesson Plan:

1. Set yourself up to iterate – start small & release often
2. Empower designers – they define success and get the business to agree with it
3. Data is power – define ways of continually getting more
4. Model-based Testing
5. Measure performance regularly – take baselines, conduct experiments & measure deltas
6. Automate, automate, automate!
7. Measure the CX “outside in”
8. Leverage API’s
9. Inspect the entire journey closely – before, during and after the chatbot
10. As true AI-based chatbots emerge, you’ll need AI to test AI
Lesson #1: Set yourself up to iterate
Assess testing effort per release using manual techniques

Manual Voicebot Tester

Manual Chatbot Tester

Manual Test Data Queries

Manual Test Data Insert

Typical Voice and Text Bot Infrastructure

Audio Transcription Service

Conversation Intelligence

Transactional Services

Mock Transactional Services

Speech To Text (STT) Technology

Dialog Manager (DM)

Statistical Semantic Interpreter (SSI)

System 1

System 2

System 3

Test Infrastructure
Lesson #2: Empower Designers

Designers do more than just design

Designers need lead from the front to achieve a high-quality Chatbots

- Previously, IVR & web technologies generally programmed customers to adapt
- Designers and linguists now control how much those constraints open up
  - Lead the discussion on what types of tasks can be completed in-chat
  - Turn a technology fashion statement into a functional solution for users
  - Expressing dialog specifications that define the envelope of possibility
  - Align to and manage user expectations and survey consumer sentiment
- Teams need educational help and mindset changes to face linguistic complexities

Consider *Design-driven Assurance* – it guides designers to specification completeness

The measure of alignment of the Customer Experience (CX), as conceived of and documented during design, with the observed CX after implementation and deployment
Lesson #3: Data is Power
To sustain a chatbot, you must always be getting more

Data Sources:

1. Testers and Business Users contribute additions
2. Generated variations using machine learning
3. Cleansed production data extracts from real-world customers
Iterative Generation of Conversational Variants

Core Test Data

Human additions

Machine-generated

Test Automation

Results Review

Curated Test Results

Bootstrap with training & reusable data sets

Each iteration enriches data set

Machine Learning
Lesson #4: Model-based Testing

- Visualization of possible interaction flows
- Collaborate across Business, Design, Dev, QA, NetOps/IT, InfoSec, Telecom, etc.
- CX Models are common
  - Define acceptable interactions
  - Capture a testing framework

- Move and size viewable area with mini-map
- Zoom in/out or fit to screen
- View model and manage design details
Lesson #4: Model-based Testing

Customer Experience Interaction Flow

• Embed quality criteria
• The Big Win: test case generation
  • Entire design is covered
  • Test cases get updated as the model evolves
With test cases generated from the model, who controls them?
Extending the Model with Customer Personas

- Extends model-based testing by simulating a variety of customers
- Test data that can be reused across many test cases
- Cover all the logic branches of complex CX flows

<table>
<thead>
<tr>
<th>Persona</th>
<th>Customer Type</th>
<th>Age</th>
<th>Customer#</th>
<th>Email</th>
<th>Mobile Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tammy</td>
<td>Gold</td>
<td>37</td>
<td>123930</td>
<td><a href="mailto:test1@xx.com">test1@xx.com</a></td>
<td>650-233-2855</td>
</tr>
<tr>
<td>Phil</td>
<td>Silver</td>
<td>28</td>
<td>694033</td>
<td><a href="mailto:test2@xx.com">test2@xx.com</a></td>
<td>510-424-4982</td>
</tr>
<tr>
<td>Larry</td>
<td>Bronze</td>
<td>54</td>
<td>384349</td>
<td><a href="mailto:test3@xx.com">test3@xx.com</a></td>
<td>617-281-2329</td>
</tr>
</tbody>
</table>
Lesson #5: Measure Performance at Every Step

- Design metrics around quantifiable definition(s) of success
- Each iteration of underlying services, application code, conversational data, etc. can improve or degrade the end user experience
- Automate for consistent measurement – it will pay off hugely
- Break down the large performance goals
  - Small, achievable sub-goals for each area are more easily digested
  - Many iterations help you realize those sub-goals in steps
Lesson #6: Automate, Automate, Automate!

- Automate test execution
- Automate your data preparation
- Automate your test case generation
- Automate with Continuous Integration
- Automate your deployment process
- Automate your performance and responsiveness measurements
- Automate your production monitoring
- Automate your processing of interactions with “feedback”, unrecognized words, and chats that escalations to the wrong type of agent
Lesson #6: Automate, Automate, Automate!

Typical Voice and Text Bot Infrastructure

Audio Transcription Service
- Speech To Text (STT) Technology

Conversational Intelligence
- Dialog Manager (DM)
- Statistical Semantic Interpreter (SSI)

Transactional Services
- System 1
- System 2
- System 3

Test Data Management Service
- Test Data

Test Infrastructure
- System 1
- System 2
- System 3
Lesson #6: Automate, Automate, Automate!

Scaling up Test Automation requires infrastructure

Execution Services
- Transcription Exercise Engine
- Chat Execution & Scoring Engine
- Test Data Management Service

Test Data Layer
- Transcribed Audio Data
- Conversation Data
- Request/Response Patterns
- Personas & Scenarios

Tools
- Transcription Preparation
- CX Model Authoring
- Service Virtualization
- Test Data Editors
Lesson #7: Measure the CX “Outside in”

- Measure as a customer would feel it and as an agent would receive it
- All that really matters is that it’s a productive use of time
  - Measure the consumer experience with synthetic customers
  - Measure the agent experience with synthetic agents
Lesson #8: Leverage API’s

• REST API & SDK-based testing is required for some aspects
  • Simulation of customers and agents at scale
  • Access to verify context data transfers between systems
  • Simulating users on FB Messenger, AMZN Alexa, Apple iMessage Business Chat
Lesson 9: Inspect the Whole Journey End-to-End

Web Automation Framework

Web Chat Automation

Virtual Chat Agents

Web Site

Chat Bot

Live Chat

- Sales Chat
- Retention Chat
- Service Chat
Lesson #10: Using AI to test AI

Developing, maintaining and executing automation requires applying AI technology:

- ✓ Speech generation (TTS)
- ✓ Speech transcription (STT)
- ✓ Natural language processing
- ✓ Computer vision
- ✓ Machine learning
- ✓ Neural networks

CX Assurance is actually a specialty area of AI
Lesson #10: Eventually, you’ll need AI to test AI

• The vast majority of chatbots aren't actually intelligent - they are built based on a decision-tree logic or a rules-based system

• Semi-intelligent bots are as smart as the designer/programmer who created it

• Bots with linguistic and natural language learning capabilities are still quite rare

• Need not have machine learning in your chatbot (yet) to get your “AI Badge”

• Humans In The Loop (HITL) classify test cases, providing a feedback loop
  • Useful Test Cases that should Pass
  • Useful Test Cases that should Fail
  • Test Case is Not Useful

• Tomorrow, machine learning let’s us build synthetic testers!
Summary – Getting Chatbots Right

• Attitudes about Chatbots will be set back 10 years if we take shortcuts
• Functional test coverage has some unique challenges
  • Designer empowerment and linguistic experience are far more critical
  • More emphasis on test data than on test scripting itself
  • Automation, automation, automation
• True adoption of AI is just beginning to emerge, and it’ll take AI to test
• Standards will be rising quickly, so tool up for many iterations ahead!
Thank You!

Questions?