Assignment M4

Saminathan Raju
sraju30@gatech.edu

Abstract—Traditionally, to get a question answered by a software program involved using a search engine or filling out a form. A chatbot allows a user to simply ask questions in the same manner that they would address a human.

In this series we're going to take a look at the chatbot used at BestBuy.com[4], (American multinational consumer electronics retailer) that is used to provide Customer Service — A chatbot can be used as an "assistant" to a live agent, increasing the agent's efficiency. When trained, they can also provide service when the call center is closed, or eventually even act as an independent agent, if desired. *Please note:* The terms "chatbot" and "virtual agent" are the same.

1. QUALITATIVE EVALUATION: From Textual Prototype

1.1 Evaluation plan

I'm going to be using the *Textual* prototype and hybrid of *Survey and Interview* to feed this portion. I'm planning to have 3 participants for this section; 2 from my family and one of my friends (I just simply asked them for a favor :-)). For one of my family members, the evaluation will take place in-person, and for the other two, via a Zoom[5] call.

The individual interaction is going to be *live* (i..e no prior experience required) and *synchronous* session with just *one interface per user*. Finally, we will be adopting a *post-event protocol* to sum it all up.

We will have *note taking* and *video recording* to record the sessions and data points discussed. Discussed with all three participants for video-recording and note-taking; they're fine with both forms.

1.2 Content of Evaluation:

1.2.1 Some of the important questions from my survey are, (Details in appendix) - Answer choices are separated by a "/"

a. Given a choice, would you like to interact with a Chatbot (aka. Virtual Agent) for customer support?

   (Yes / No / Maybe)

b. How often would you like to interact with a Virtual Agent as opposed to a live agent?

   (Always / 1% - 25% of the time / 26% - 50% of the time / 51% - 75% of the time / Over 75% of the time / Never / Not applicable (I don't visit the store, or website at all)

c. What are your primary expectations from the Virtual Agent?

   (Relevance of Results / Quicker Turnaround Time / Answering to the point / Ability to complete my requests fully / Transferring to a Live Agent with my feedback and chat history / Answering FAQs)

d. How important is it for a Virtual Agent to process voice commands? (i.e. When you're on the move, Driving a car)

   (Important / Not important)

1.2.2 Some of my interview questions are, (Details in appendix)

a. What device do you usually use to chat?
b. What are the usual challenges faced?
c. What capabilities do you think the chatbot should have?
d. How would you like chatbots to handle sensitive information?
e. From the Warehouse user point of view, what kinds of manual data collection the chatbot can help with?

Please note: the interview will be semi-structured. There will be follow-up questions and the interviewees will just be told what to do, not how.

1.3 Goal

The goal of this assessment is to validate the overall goal of the interface[2], and to check if the initial set of requirements still holds good. Even though the survey is asynchronous, having the data point ahead of time will help greatly to drive this session.

1.4 Discussion

The users are going to be at different locations and operate with the chatbots in a variety of contexts ranging from orders to products, accounts, payments etc. One simple goal they have is "***relevance***": the chatbot should be able to perform the relevant activity. The device they need will range from a simple mobile device to laptops; since we're talking about enabling voice commands, we will also extend this to any smart assistants. The only physical thing the user needs to do is "triggering the correct utterance(s)"; there's very little cognitive load on the user as the bot is equipped with a "help" mechanism based on the current user context; hence, absolutely no learning curve is required.

This evaluation will help me understand if my requirements were defined correctly. It should be *specific, evaluable* and *functionling*: We have seen it already - the difference between live agent answering, vs the bot answering needs to be fairly subtle. *Usability* is another key requirement we expect this to meet - understand the channel of request and respond via the same. No learning curve and voice assistance (using JAWS[3] when needed) should ensure the *learnability* and *accessibility* part is covered. Most importantly, this evaluation should help me in these three surprise areas, I call it 3Cs: *Cost, Compatibility*, and *Compliance*. Cost factor is technical and I don't expect a normal user to be aware of that - for that reason, I have my friend who's technically knowledgeable as one of my participants. Compatibility is technical too to some extent, but the hardware / device manufacturer takes care of timely updates, so the cognitive load on that area is greatly off-loaded from the user.

2. EMPIRICAL EVALUATION: From Verbal Prototype

Our goal here is to have verifiable and conclusive data on how to build the interface. As a part of this evaluation strategy, I will be creating two groups with about 30 participants each. Each of them will participate in both the treatments (described below), so ours is a "***Within subject design***". This helps us get twice as much data, and to capture subtle effects on each individual's perception about our interfaces (described below).

Randomization: In order to help control biases, the participant allocation will be randomized between groups.

Our treatments: Since ours is a new task, we will be making subtle variations in our interface designs and creating two treatments for comparison.

While looking at the objective (based on Survey, and M3[1]), three important things the user expects out of the system are: ***Less turnaround time***, ***Improved Speed***, ***good performance***. They all come to "nominal" value (as measured in "seconds") - so a **"chi-squared test"** is used; we're looking at a specific critical item here; also comparison with a familiar entity (i.e. store agent) will help seamless correspondence.

- ***Control Variables***: Variables that do not cause the difference in treatments
    - "Room lighting, Room temperature, size of the display, mouse positioning, keyboard angle, chair height and so on"
- ***Environment Variables***:
    - Background noise can trigger improper utterance (in-case of voice commands, also traffic noise while on the move)
- ***Experimental Variables***:
    - *Independent variable*: think of this scenario: "getting complex queries solved by the chatbot vs live agent"
        - *Treatment 1 - Interface A*: Chatbot handling everything; may take some time and ask follow ups, but will help out with the request.

        - *Treatment 2 - Interface B*: Chatbot gathering basic details and transferring to live agent with chat history and request
    - *Dependent variable*: Turnaround time across user utterances
- ***Null Hypothesis***: Turnaround time of Interface B is not quicker than Interface A (OR Interface A and B have the same turnaround time)
- ***Alternate Hypothesis***: Turnaround time of Interface B is quicker than Interface A
- ***Completion***: The participants will go through the verbal prototype flow for different treatments. The data will be recorded in the form of "note-taking" which is analyzable.
- ***Data generated***: Distribution of turnaround time information between the two interfaces
- ***Analysis on the data***: To help us understand whether the difference is too much (> 5% variance) to consider alternate hypotheses a better one.
- ***Hidden variables to mess-up with the data***: Wake word (like Hey Siri, Okay Google, or "alexa") and background noise - especially when the user has more than one device in the environment that are operated by others, Or background noise especially on the move can misinterpret the results.

3. PREDICTIVE EVALUATION: From Wizard of Oz prototype

This is an evaluation method that can be done without the users, maybe an expert, Or the designer himself (which is me) can perform this evaluation. As a part of the rapid feedback process, we are going to adopt this. I chose Wizard of Oz because of the widespread scope of *Heuristic Evaluation*.

We are going to be performing "***Cognitive Walkthrough***" task analysis, by simulating the user's mental model at each stage. We can start by constructing specific tasks that can be completed within the prototype.

3.1 Goal: To be able to use the chatbot with full relevance.

3.2 Cognitive Walkthrough: (Given only a subset of a flow. Please refer appendix for the complete conversation)

Launching the Best Buy chatbot is a straightforward process by going to www.BestBuy.com → Menu → Support & Services → Visit our Support Center. Soon after launching the chatbot, it offers a greeting message along with the help text. I can either build from the help text, start typing my own, or tap on the mic icon to speak.

"Me (Pretending to be the user): Hey BB, Can you please list all Samsung 40" TVs?

BB: Sure. But the list is too big to read out. Do you want to further filter it? You can say things like, Ratings over 4* and Android TVs, Or HDR+ ready.

Me: Over 4* and HDR+ ready

BB: Great. I have found 3 items. Do you want to further filter, Or hear it out?

Case : Further Filter	**Case : Hear it out**
Me: Further filter by price under $400. BB: Sure. I found 1 item. What do you want to do with it? You can Add to cart, Buy now, Details, Cancel, Or Help.	Me: Hear it out BB: The first item is Samsung Smart HDR+ 40" for $449. You can say Add to Cart, Buy now, Details, Next (or position or last), help, further filter, Or Cancel
Me: Help BB: Sure. You are in the context of a product catalog. In addition to keywords like I said, you can say it in your own way. For instance, *"add this to cart and checkout using my saved details"* will work just fine.	Me: Last BB: The last item is Samsung Smart HDR+ Bazel free 2021 edition at $549. You can say Previous (or position) or First, in addition to other options. "

Per the conversation above, it is clear that the ***gulf of execution and the gulf of evaluation is almost non-existent***. That is the power of a data-driven approach. It helps us eliminate all surprises once could potentially get enroute.

A Thing to improve: providing buttons on the interface (yellow highlighted below), in addition to other modes of input. It will be useful when the user wants to provide an input with a click as opposed to typing lengthy utterances.

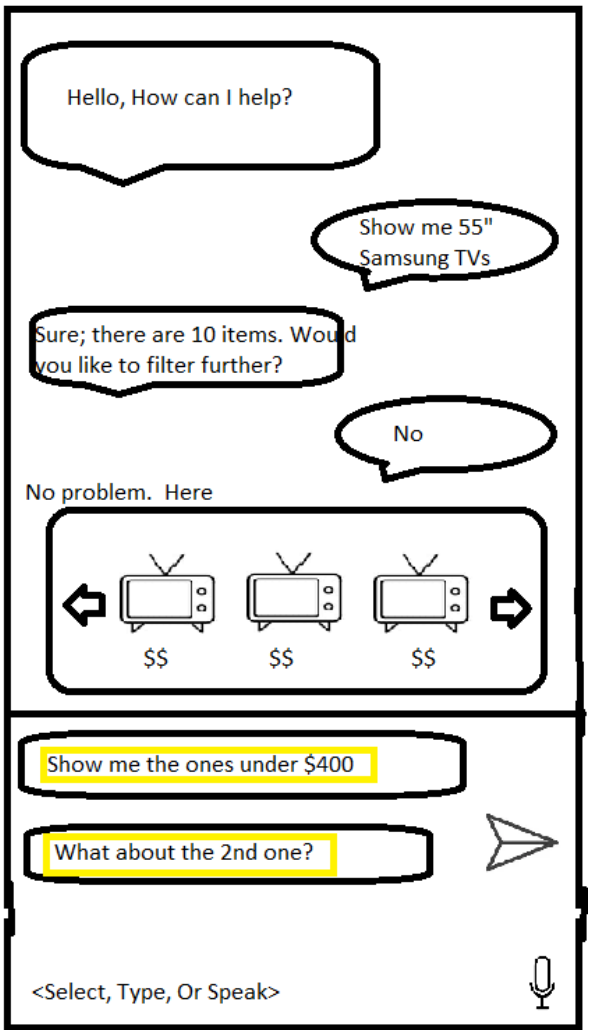

Figure 1*: Only for an idea; not the actual / proposed model*

Because the evaluation is done without the users, there's a tendency to agree to what the interface offers as we are the ones who incubated it. However, I tried my best to think from the user's perspective and provided data points. It is possible to continue the conversation at any time either by voice, or text (either by typing, Or selecting the options from the pre-constructed ones (like highlighted above))

As I explained in my prototype[1], "*help*" is the keyword to provide users with contextual help wherever they are in the conversation; also, soon after logging in, it is the bot that greets the user with a welcome message, also shows the user the sample utterances as an example. With all the UI cues, one can just start using the system without a learning curve. Thanks to the Natural Language Understanding (NLU) engine that helps with the understanding of nuance and flow.

4. PREPARING TO EXECUTE

4.1 Predictive Evaluation:

I chose this because of Heuristic Evaluation. It gives a snapshot of what users might think about the interface and what the practical challenges could be. I understand that it is not evaluated with the actual user, but I consider this as an opportunity to evaluate against the individual brainstorming exercise (which is considered one of the most effective ways of brainstorming) which was performed at the initial stages. I come from an Artificial Intelligent background and developed one such tool - so one other reason I chose this approach is to have an opportunity for "Simulation based evaluation" in future.

4.2 Empirical Evaluation:

When the goal is providing relevant results with optimum performance, there can be countless possibilities and improvisations one can make to the interface. I chose "Empirical Evaluation" because of its flexibility to choose from different treatments. There's no right or wrong decisions here as long as the goal ends up in "user centered design".

One other thing I like about this evaluation is closed ended decisions are made based on hypothesis. We're not debating on a range, but rather focussed on a value (which is turnaround time in this case). This clears the clutter and helps with the decision making.

REFERENCES

1. My M3 submission
2. My M2 submission
3. https://www.freedomscientific.com/products/software/jaws/
4. https://en.wikipedia.org/wiki/Best_Buy
5. https://en.wikipedia.org/wiki/Zoom_Video_Communications

APPENDICES

- **Abstract - Deep Dive**

  The use of chatbots evolved rapidly in numerous fields in recent years, including Marketing, Supporting Systems, Education, Health Care, Cultural Heritage, and Entertainment. Chatbots, or conversational interfaces as they are also known, present a new way for individuals to interact with computer systems. You should think of a chatbot when you have

  - Global presence, growing customer base, with multi-lingual support
  - To provide channel agnostic support
  - To reduce cost by providing 24 x7 automated support
  - To scale on demand without employing additional headcounts

- **Link** to Survey Analytics
- **Links to Interviews**
  - A customer's view: https://youtu.be/BsUPiPveZbg
  - A store associate's view: https://youtu.be/rUwVLT4Eu5c
  - An analyst's view: https://youtu.be/hBHH013ANts
  - **More key interview questions:**
    1. Can chatbot alert the user in advance?
    2. Can chatbot be used in historical data analysis?
    3. What kind of predictive analysis can the chatbot help the user with?

4. Can chatbot replace live agents with adequate training?

- **Predictive Evaluation: Complete conversation (one full cycle)**

Me: Hey BB, Can you please list all Samsung 40" TVs?

BB: Sure. But the list is too big to read out. Do you want to further filter it? You can say things like, Ratings over 4* and Android TVs, Or HDR+ ready.

Me: Over 4* and HDR+ ready

BB: Great. I have found 3 items. Do you want to further filter, Or hear it out?

Case : Further Filter	**Case : Hear it out**
Me: Further filter by price under $400. BB: Sure. I found 1 item. What do you want to do with it? You can Add to cart, Buy now, Details, Cancel, Or Help.	Me: Hear it out BB: The first item is Samsung Smart HDR+ 40" for $449. You can say Add to Cart, Buy now, Details, Next (or position or last), help, further filter, Or Cancel
Me: Help BB: Sure. You are in the context of a product catalog. In addition to keywords like I said, you can say it in your own way. For instance, *"add this to cart and checkout using my saved details"* will work just fine.	Me: Last BB: The last item is Samsung Smart HDR+ Bazel free 2021 edition at $549. You can say Previous (or position) or First in addition to other options.

Case 1: Me: 'Cancel':

Case 1: BB: Sure. How can I help you now?

Case 2: Me: 'Details':

Case 2: BB: This is the least expensive Samsung Smart TV you can buy today! <and a few other important details>. You can say, Add to Cart, Cancel, Buy now, go back, or Help

Case 3: Me: 'Add to Cart':

Case 3: BB: Added to cart successfully. Would you like a 5 year unconditional warranty for an additional $49?

Case 'Yes'	Case 'No'
BB: Thanks. Added successfully. Should I check the item out, Or continue shopping?	BB: No problem. Should I check the item out, Or continue shopping?

Case 4: Me: 'Buy now':

Case 4: BB: This item arrives tomorrow (Or the date) to your default shipping address. Shall I place the order using your default payment mode?

Case 'Yes'	Case 'No
BB: Thanks! The order was placed successfully. What would you like to do now?	BB: Unfortunately, this is not the correct channel to add new payment information. Would you like to pick up from your nearest store at "500 N.Pines Blvd, Hollywood 33025" by tomorrow evening? <Yes / No conversations after that>

Case 5: Me: 'Continue Shopping'

Case 5: Sure. How can I help you now?

Case 6: Me: 'go back' or 'Back'

Case 6: That was the last conversation in my memory. How can I help you now?

- BB → Acronym I gave for Best Buy's chatbot; 'Me' → User / customer interacting with the chatbot