

Usability Test Report:

Adam's Portfolio Chatbot Prototype

Date of Report: January 20, 2021
Date of Test: January 2021
Location of Test: Remote

Prepared for: The Digital Assistant Academy
Email: info@digitalassistant.academy

Prepared by: Adam Woollacott
Email: amwoollacott@gmail.com

Background Summary

This usability test was completed to determine the effectiveness and usability of the Portfolio Chatbot Prototype. It took place remotely with 5 participants, via zoom allowing them to screenshare with the test facilitator who was watching live. The users were each sent a link to the test, which were conducted through Botsociety's chatbot prototyping web application. Each user was given tasks to complete to ensure the chatbot was thoroughly tested. After each test, the participants filled out a specific questionnaire that was designed to measure the usability of the chatbot, - CUQ, Chatbot Usability Questionnaire - that calculated a score out of 100.

The aim of this test:

1. To find out what paths work and don't work within the bot.
2. Develop a deeper understanding of how to improve the bot.
3. Increase user data of the words and utterances.

The Portfolio Chatbot performed well overall and scored 77.8+-16.1 standard deviation. However, the participants did encounter a number of problems such as the speed of the messages and the amount of text and information were delivered. The tests also showed that buttons were a preferred method of input, rather than typing, all of which will be discussed in this report.

Methodology

Participants tested

Five participants, having the following characteristics, evaluated Adam's Portfolio Chatbot prototype.

Age

18-25	0
26-39	3
40-59	0
60-74	2
TOTAL (participants)	5

Gender

Women	2
Men	3
TOTAL (participants)	5

Occupation

Store Manager
Business Owner
Occupational Therapist
Project Manager
Recruitment Consultant

What the participants did:

After the participants were chosen, they each received an email that included documents that allowed them to read the information about the test and what was expected from them throughout. They also received and were asked to sign a consent form to ensure that they had read understood what the data will be used for. Within the emails they received links to:

- Zoom meeting.
- The chatbot test platform: Botsociety.
- Questionnaire.

All the participants were given an individual time slot, and once in the Zoom meeting, they were introduced to the test facilitator, who briefly ran through what was expected and the instructions for the test. Each participant had 30 minutes to go through a number of situations that were provided. They were each asked to think aloud whilst interacting with the bot, explaining their reasoning for choosing options, including any positive or negative opinions. This method; Concurrent think aloud (CTA) was chosen ahead of other methods as its was useful to get real-time feedback and emotional responses from the participants before they forget.

At the end of the test and allotted time, the participants were asked to share their overall thoughts with the test facilitator before ending the Zoom call and submitting the questionnaire.

What data was collected:

Each participant was asked to complete the same set of tasks:

1. Find information about the bot so you get a better understanding then find Adams Bio.
2. Find out what Voice Interaction Design is via education, and then Contact him.
3. Look through Adam's Hard and Soft skillsets.
4. Try to see if you can view the projects he has been involved in throughout his experiences.
5. Self-ran search - Go through other paths of your choice.

The idea to create a self-ran search was allow the user to navigate themselves around the bot without any preconceived ideas or additional thoughts. This task was to recreate how users would normally communicate with the bot in a natural circumstance.

The participants each screenshared their screen so that the facilitator was able to follow along throughout the test. Notes were also taken of any of the participants comments alongside any errors within the system. The whole test was recorded for review purposes after the test had been completed.

At the end of the test, the participants completed an online questionnaire that included a Chatbot usability questionnaire (CUQ). This was validated as part of a Research PhD at Ulster University [1] and chosen as part of this test due to its adaptability with testing different systems.

The questionnaire consisted of 16 questions that related to different areas of the chatbot to measure its usability. Half of the questions focused on positive area's and the other half on negative areas. Each of the questions could be answered on a five-point scale from 'Strongly Disagree' to 'Strongly Agree', with the participant choosing which relates best to each statement.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The chatbot's personality was realistic and engaging	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot seemed too robotic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The chatbot was welcoming during initial setup	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 1: Screenshot of selection of Questionnaire

Test Results

The results from the CUQ were calculated using the CUQ calculation tool created by The Ulster University PhD team [1]. The scores were calculated out of 100, to ensure accurate comparisons with the System Usability Scale (SUS). The overall score of the Portfolio chatbot was 77.8, which was above the average industry score of 68 [2]. The lowest score was 51.6 by participant 3 and the highest was 93.8 by participant 2.

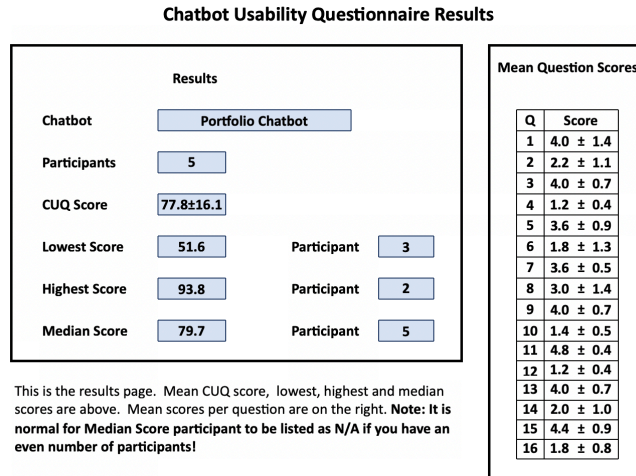


Figure 2: Chatbot Usability Questionnaire Results

According to the results from the questionnaire [Figure 3], 80% of the participants were satisfied with the bot prototype, with one participant unsure. However, Figure 4 shows that 40% of the participants found that the conversation was not overly easy.

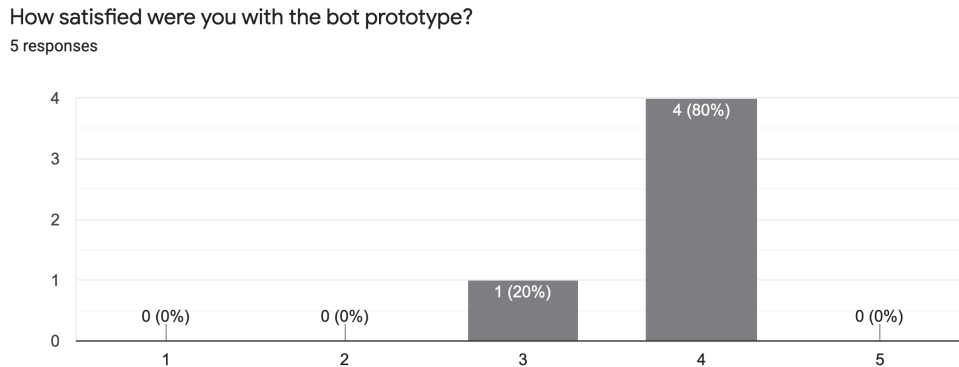


Figure 3: Chatbot Satisfaction Question.

Did you find the conversation between yourself and the bot easy?

5 responses

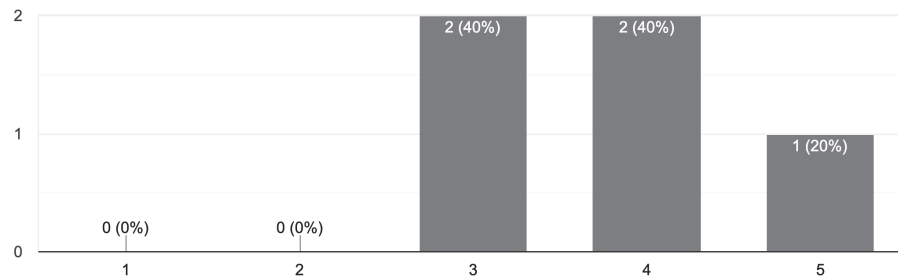


Figure 4: Chatbot - How easy was the conversation?

The overall findings and recommendations from the participants were as follows:

1. The messages were too fast, couldn't read them quick enough.
2. Unable to scroll up or down to view previous messages.
3. A few minor path errors and inactive buttons.
4. Messages were too long, must be more concise and direct.
5. Fun, engaging idea.
6. Users preferred clicking buttons rather than typing.
7. Too many options to view, bit overwhelming/confusing to choose.
8. Potentially a little long for the user, could be easier via other methods, however the concept is niche and it's fun to use.
9. Reduce unnecessary words that don't add much to the conversation.
10. Liked the hyperlinks to the portfolio.

Discussion and Analysis

The results in this report tend to show a reasonable score of usability throughout, but by comparing the results to the participants comments, shows that the bot has lots of room for improvement.

Findings 1 and 2 both cover similar questioning regarding the ability to be able to see multiple or previous messages after they have been moved off the screen. As the output platform of the chatbot being Facebook Messenger, they do incorporate scrolling into the platform. However, this function could not yet be activated due to it being at a prototype stage of the design.

With the chatbot including a variety of buttons and typing responses, 3 out of the 5 participants preferred to click on buttons, rather than to type answers in. They became confused as to what to do, which resulted in them having to read the instructions more than once, on numerous

occasions. In order to solve this, the chatbot could be altered to mainly use buttons, therefore keeping typed responses to a minimum and ensure that it is clear to the users.

Throughout the testing, the users found a number of errors that led to unusual paths or 'dead-ends'. Whilst normally, errors are a failure of the bot, throughout this testing phase, errors are welcomed, as it shows that the bot is getting thoroughly tested in prototype stage. By knowing the errors at this stage allows for corrections to be made before getting released. Some of these errors included:

- An inactive 'Audio Editor' button.
- The secondary 'Experiences' menu doesn't lead anywhere.
- 'Audio Restoration' path drove users to another path.
- Yes/No answers – both answers went to the same path.

Further corrections would be made in the future to establish fully functional conversation paths, before embarking on another round of testing.

Findings 4 and 9 suggest that there are too many words and too much irrelevant information given to the user. This particular chatbot was created to be informative, but sometimes there is a fine line between the right amount of information and too much information. 60% of the participants would have preferred shorter and more concise information. This would result in a slight reduction of information given but given in more readable and bite sized messages. Participant 4 also suggested utilizing bullet points to help to keep information concise. Whilst it feels a less conversational style, they would help to increase the readability of the information provided. For future use, longer messages would be rewritten to reduce the amount of words and will include bullet points where possible.

One participant felt it was unnecessary to include words and phrases that did not relate to the information. They preferred a more direct approach. Additional words or phrases such as '*Perfect*', '*Alright, no worries*' and '*Nice one, I'll try and give you brief explanation*'. These were initially included to add a more conversational approach and personality to the chatbot. Moving forward, alterations would be made to ensure more direct messages are included, whilst also keeping a certain 'personality' to the chatbot. Another comment was raised with the wording of some of the buttons. Words like 'Hard Skills' and 'Soft Skills' were questioned and misunderstood. Perhaps these words were too focused on resume writing and should be broadened to ensure it's understood by all types of user.

From looking at the data provided from the participants when entering written answers, the utterances provided within the bot, covered most of the options that the participants entered. The new utterances will be added into the bot and constantly updated throughout further testing to ensure a wider selection. However, what was interesting to find was the option to return to the main menu at any point throughout the bot. A number of times, the participants decided not to carry on down a certain path and wanted to return to the main menu. Instead, they had to carry down the path until the option was given to them. This relates directly to the next point below.

A number of the participants commented that the chatbot was a bit laborious. Although they were naturally taking longer, due to them thinking out loud for the purpose of the test and some system errors, the participants believed that it took too long to find information. With people's attention span on the decrease [3], it is important that the chatbot ensures that it captures people's attention, or at least allow them to find information about Adam's portfolio within a shorter timeframe. There are a number of ways that could help these issues; The first being to reduce some of the options available to the user, allowing them to find the bot a little simpler to understand. The second option would be to incorporate Natural Language Processing (NLP) into the design. Currently the chatbot is mostly a flow-based chatbot at this current stage. This method has both pros and cons. For the former, it is clear for the user to know what they can

achieve with the bot guiding them through the conversation. However, for the latter it can be very restrictive by not allowing the user to change the topic of conversation quickly.

What is worth noting however, that during this prototype testing phase, NLP was unavailable through the platform. The next stage would be sending the data to an integrated platform such as Dialogflow to continue the design using dialogue management and natural language processing. The aim would be to be able to merge the flow-based with natural language to further increase the chatbots usability and power.

Conclusion

In conclusion, the participants were reasonably satisfied with the Portfolio Chatbot Prototype, finding it fun and interesting to use. There were a few concerns about the length of time it took to move around the chatbot, but with the alterations and revisions indicated in this report, along with the addition of some natural language processing, it is hoped that this will significantly reduce the time spent to find information. Further testing would be needed after these alterations, to ensure that improvements are made. The inclusion of more utterances would constantly be monitored to improve the machine learning capabilities.

References

1. Samuel Holmes, Anne Moorhead, Raymond Bond, Huiru Zheng, Vivien Coates, and Michael Mctear. 2019. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. In Proceedings of the 31st European Conference on Cognitive Ergonomics (ECCE 2019), Maurice Mulvenna and Raymond Bond (Eds.). ACM, New York, NY, USA, 207-214.
DOI: <https://doi.org/10.1145/3335082.3335094>
2. Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2018. Assessing the usability of a chatbot for mental health care.
3. Jacel Booth. 2020. Visions for 2020: Key trends shaping the digital marketing landscape. [Web] <https://blogs.oracle.com/oracledatacloud/2020-trends-shaping-the-digital-marketing-landscape>