American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
http://asrjetsjournal.org/

Chatbot Quality Assurance Using RPA

Ganesh Gatla*

480 Meadow Hill Drive, Alpharetta, GA,30004
Email: ganeshgatla73@gmail.com

Abstract

Chatbots are becoming mainstream consumer engagement tools, and well-developed chatbots are already transforming user experience and personalization. Chatbot Quality Assurance (QA) is an essential part of the development and deployment process, regardless of whether it's conducted by one entity (business) or two (developers and business), to ensure ideal results. Robotic Process Automation (RPA) can be explored as a potential facilitator to improve, augment, streamline, or optimize chatbot QA. RPA is ideally suited for tasks that can be clearly defined (rule-based) and are repeating in nature. This limits its ability to become an all-encompassing technology for chatbot QA testing, but it can still be useful in replacing part of the manual QA testing of chatbots. Chatbot QA is a complex domain in its own right and has its own challenges, including the lack of streamlined/standardized testing protocols and quality measures, though traits like intent recognition, responsiveness, conversational flow, etc., are usually tested, especially at the end-user testing phase. RPA can be useful in certain areas of chatbot QA, including its ability to increase the sample size for training and testing datasets, generating input variations, splitting testing/conversation data sets, testing for typo resiliency, etc. The general rule is that the easier a testing process is to clearly define and set rules for, the better it's a candidate for RPA-based testing. This naturally increases the lean towards technical testing and makes it moderately unfeasible as an end-user testing alternative. It has the potential to optimize chatbot QA in conjunction with AI and ML testing tools.

Keywords: Chatbo; Quality Assurance; RPA.

1. Introduction

Chatbots have become a vital consumer engagement tool for businesses. Rudimentary chatbots rely upon automation, whereas modern chatbots lean more heavily on Artificial Intelligence (AI) and Machine Learning (ML). Many chatbots are already integrating Natural Language Processing to facilitate their target audience better (through voice chat). Custom chats offer more versatility but out-of-the-box chatbot applications, which can be repurposed for a variety of industries and engagement goals, are also commonplace.

----

* Corresponding author.

This variety and constant evolution cycles chatbots are going through make creating a standard set of Quality Assurance (QA) protocols challenging and, in some cases, impractical. However, creating a QA procedure from scratch for every chatbot an organization uses is a resource and time-consuming. This paper aims to identify Quality Assurance practices common for most varieties of chatbots and whether or not these QA practices can be automated through Robotic Process Automation. The quality assurance process for chatbots starts with defining the necessary and secondary dimensions [1], including integration, interaction, and analytics. The primary testing parameters include identifying intent, interpreting complex requests, interpreting multiple requests in one message, quality of advice provided [2], the hierarchy of advice provided, escalation system to human agents, UX/UI, safety, relatability, etc. The parameters are slightly different for different industries and levels of chatbot sophistication. Apart from the encompassing dimensions mentioned above, there are a few established "Service Quality Models" [3] that can be applied to chatbots, but they are broad spectrum and are used to evaluate the quality of service of websites and mobile applications as well. Therefore, it makes them useful for evaluation from an information delivery perspective. Another set of quality dimensions that is more macro than specific areas of exploration/QA discussed above are Understandability, reliability, assurance, and interactivity [4]. These can be further broken down to conduct specific Quality Assurance tests. One of the core challenges associated with chatbot QA is the complexities of human-machine interaction when Natural Language Processing is still evolving. Robotic Process Automation (RPA) has evolved and improved over the years, but it has certain limitations which are becoming more pronounced with advances in Artificial Intelligence (AI). However, the two different technological approaches can coexist and may prove useful in a broad array of applications, including chatbot QA, but this is not included in the scope of this paper.

In order to understand how RPA can be applied to chatbot QA, we will first go over the defining traits [5] of an RPA. Repeatability is at the core of an RPA, alongside properly defined tasks. Unlike self-learning Artificial Intelligence models, an RPA-based process is heavily reliant on a user's ability to define certain tasks. For chatbot QA, these tasks would be defined by the tester. If it takes more time to define and implement RPA procedures than to manually conduct a chatbot's QA, the RPA will be redundant unless it can be repurposed (at minimal effort and time cost) for the QA of a wider variety of chatbots. The effectiveness of a set of RPA procedures will be determined by how many different chatbot variants it can be applied to, taking the modifications/adjustments for each chatbot into account.

The paper aims to determine if any chatbot QA tasks can be automated following the best RPA practices. If yes, are they efficient enough to be worth the time and effort of the individuals/resources involved?

2. Chatbot Quality Assurance

The first chatbot Eliza was created in 1966. It was rudimentary and could only maintain limited conversations. The first AI-powered chatbot (Jabberwacky) was built in 1988. But a chatbot that could help users/individuals with certain tasks online and provide relevant information wasn't introduced till 2001. Then came Apple's Siri, a mainstream application that combined chatbots with Natural Language Processing (NLP) [6]. Chatbots are being extensively used in retail, customer service, and online help desks, with varying levels of sophistication.

The Quality Assurance (QA) process for chatbots varies based on a number of factors. And for a comprehensive evaluation, the QA process may be broken down into two main segments: Understanding and response.

For the understanding segment, a chatbot may be evaluated for its ability to recognize intent [7]. Identifying the intent can trigger the right response, increasing the probability of keeping the user engaged in conversation. There are other QA parameters, both separate and associated with intent, starting with coherence, which tests whether a chatbot can discern between coherent and noncoherent messages. Then there are QA tests for sturdiness to ensure that the chatbot can understand the message despite the presence of typos and identify intent by cutting through irrelevant content. The precision QA tests determine how accurately a chatbot identifies intents [8].

For responses, the response quality of a chatbot can be evaluated based on [9]:

1. Informativeness
2. Conversation flow/fluency
3. Human likeness/Natural conversations

As the conversation progresses, context becomes an important quality that needs to be evaluated to ensure that the chatbot can carry a coherent conversation for a relatively long time.

These QA protocols are industry/use-case agnostic and can be used to determine the baseline quality of any chatbot. It can be expanded upon based on specific industrial needs. Some chatbot QA tests may also include testing for safety, response time, and identification of human/machine users.

Chatbots that act as Natural Language Interfaces (NLI), part of the next generation of User Experience (UX), still lack in response quality, the ability to hold to natural interaction conventions, and are unable to preserve context [10]. It's one of the several recognized chatbot QA challenges. Other challenges include:

1. Lack of static QA mechanisms [7].
2. Lack of clustering tools for chatbot QA, which leads to a lack of collective repositories of utterances and testing phrases (for similar topics).
3. No classification standards for chatbots for unified/translatable QA practices.

3. Robotic Process Automation (RPA)

The basic aim of an RPA is to ease the workload of humans (employees) by automating repetitive tasks [11]. This leads to higher efficiency, employee output, and fewer human errors. Gartner defines RPA software as tools that perform "if, then, else" statements on structured data, identifying rudimentary logic and the presence of a well-defined model (simple rules and business logic [12]) as the core tenets of an RPA.

RPA faces a wide array of adaptation challenges [13]. The particular set of challenges varies between organizations and industries. The challenges relevant from the context of Chatbot QA automation are:

1. Support for RPA. An organization may not prioritize Chatbot QA through RPA for a variety of financial, technical, and operational reasons, including (but not limited to) the existing technical expertise and level of chatbot sophistication.
2. Availability of potentially better technologies and tools. AI and ML-based QA tools for chatbots might yield better ROI.
3. Evaluation of benefits offered by the RPA. An incomplete set of metrics to evaluate the effectiveness of RPA for chatbot QA may prevent its adoption.
4. The complexity associated with structuring the QA processes for successful RPA.
5. Lack of expertise associated with the implementation and development of RPA bots.

These challenges will determine whether or not to use RPA for chatbot QA in general or for specific use cases.

Despite its challenges, RPA is being extensively employed for testing in certain industries, predominantly auditing. Three characteristics of auditing tasks that can be automated using RPA can be translated for chatbot QA as well. These characteristics are [14]:

1. Well-defined process/tasks (Minimal ambiguity).
2. High-volume repeated tasks (If the process has to be redefined multiple times for automation, RPA might add to the time cost instead of reducing it).
3. Mature/Well-known tasks should be automated. Not anomalies.

Versatile RPA tools can be used to test various stages of any software development and deployment (including chatbots), which may make the last stage of QA testing more efficient [15].

RPA can be used to test the quality of a chatbot *while* it’s being built all the way to the end when it’s deployed (to be tested for user interactions).

4. Chatbot QA Using RPA

The chatbot QA for an organization will be different if they are creating/coding one in-house or if they are getting a pre-made QA and deploying it. End-user testing will be necessary in both cases, and there are multiple ways this testing can be automated using RPA.

Testing for input variations. If you have variations of one question, you can define a simple RPA process to identify how many of these variations a chatbot *doesn’t* understand. For example, a user might be asking whether they are talking to a human or a chatbot. They may ask: "Are you a human?" or "Human/bot?" or even an emoji of the robot with a question mark. If you can obtain such a set of instructions, an RPA model can easily be used to catalog for which of the inputs a chatbot is not recognizing the intent [16].

An RPA can be used to create conversation flows from existing human conversation examples. There are multiple ways to define these creations, but if you set the right rules, it might be possible to create a larger testing data set from a seemingly smaller one.

A common chatbot testing technique is splitting conversation datasets into two segments: Training and testing [17], following a predefined split (like 70/30). An RPA process can be defined for creating multiple variations of these two data sets, essentially growing your training and testing data. This can also ensure that all the conversations are used for testing *and* training.

AI planning practices can be applied to efficiently break down the chatbot testing process (for QA) for automation [18]. This can be especially helpful for scenarios in which breaking down the testing process/QA process for a particular chatbot into well-defined roles is difficult or no precedent exists.

An RPA can be designed to modify a specific input, like an email address or a specific business query, into multiple variations or include typos, sometimes in the same conversation flow, to identify which variation is not recognized by the chatbot.

RPA can also be used to determine the speed of various responses to find patterns that may need correction. Speed of response is important to keep conversations going/flowing for most industries, including social counseling [19].

Another aspect of chatbot QA/testing is evaluating its ability to identify, classify (if needed), and respond appropriately to different aspects of a normal conversation (following the natural flow of a dialogue) it aims to mimic for any given language. Since it can be the same model for different types of chatbots engaging in different conversations (money changer, boarding passes, etc.), an RPA protocol can be designed for completely automated testing or human-assisted expert testing [20].

Chatbot testing/QA can be classified into two categories – technical and socio-technical [21]. The testing of technical aspects is easier to automate. The level of sophistication of a chatbot significantly contributes to the level of automation that can be employed for its testing/QA. Simply informative chatbots that are only expected to match the query to the right piece of information can be the easiest to test. Chatbots that both store user information and need to use it in the conversational flow may not facilitate the same depth of testing via RPA [22].

One aspect of a chatbot’s technical QA is the testing of its security features. A chatbot is not only connected to a business database, a connection that can be exploited by a planned attack, but many chatbots also gather user information/data, and any vulnerability in collecting, storing, transferring, and processing that data can lead to leaks which may jeopardize the user's privacy and the business' reputation [23]. Even rudimentary and easily automated security testing can at least help a business identify which publicly available chatbots have gaping security vulnerabilities, so such testing should become part of the RPA-powered part of the QA, ideally with necessary additions. The vulnerabilities of these chatbots (if not identified during QA, with or without RPA involvement) can significantly undermine the user privacy measures imposed by regulatory bodies on internet-based communication [24].

Mature testing procedures like System Usability Scale (SUS) and the relatively new Chatbot Usability Questionnaire (CUQ), if integrated into the chatbot QA, may offer more RPA usage opportunities compared to unique testing procedures with limited predictability. These have been found useful for the evaluation of a multi-language chatbot [25].

Based on an Arabic dialect chatbot that was powered by ML and was tested on multiple corpora, it may be possible to train and then test a chatbot using a specific body of text, like an SOP, operational manual, or even a catalog, to associate the query with a specific part of the text in its database. Training and testing a chatbot using a specific body of text can be automated by generating test queries based on a number of rules (and combining those rules to create complex conversational scenarios to test database matching). This may become a testing practice when adding to a chatbot’s database [26].

The chatbots designed and implemented for specific research applications might be easier to test via RPA than chatbots designed to directly engage the consumers. The reason is that these chatbots are usually designed to perform specific, well-defined tasks, which may facilitate easier RPA-based testing [27].

When leveraging a dataset like The Stanford Question Answering Dataset (SQuAD) to test a chatbot, an example of which is a chatbot framework tested for COVID [28], an RPA can be leveraged to choose and present the queries to the chatbot and develop a conversation flow.

5. Results

A few things stand out from the research papers and the studies associated with chatbot testing processes (part of the QA) and their potential for RPA.

By definition, RPA is ideal for repeated/repeatable and well-defined tasks, which makes it ideal for some standard testing processes (generic queries, variations of the same query following a formula, etc.). However, a chatbot's ultimate job is to engage a human in conversation.

Not every aspect of a human conversation can be predicted and planned for, and thus, it cannot be defined for a software robot (for RPA). As a result, its testing cannot be automated with RPA alone (maybe with the help of AI and ML).

But this also ties into the “scope” of the conversation. The broader the scope of the conversation is, the more difficult it may become to automate its testing via RPA. But the narrower the scope, the easier it becomes to predict the flow of *most* conversations. For example, a chatbot designed to take food orders for a restaurant with a massive menu may experience so many unique queries that it might become impossible to develop an RPA or a set of RPAs to test it for all eventualities, and it might just be easier to test for them manually instead of developing RPAs to cover them all.

However, the narrower the scope, the better from an RPA-based QA perspective. An automotive parts manufacturer (a B2B business) may receive 200 queries a day and over 65% of them are regarding part availability. The queries will contain either the part number or part name (or their variations). A lot of field testing of such a chatbot can be handled by an RPA.

The scope of RPA testing differs when the role of a chatbot is different from engaging with a target market. Chatbots developed for research applications might be easier to test using RPA due to the well-defined scope and parameters and the quality of testing data input. It's difficult for an RPA-based testing framework or system to completely or significantly replace the human/end-user testing of a chatbot.

It's most effective in the technical testing aspects of the chatbot QA but the closer the testing gets to the end-user use cases, the less effective RPA-based testing might become, especially if the chatbot scope is quite broad. However, if used in conjunction with AI and ML test practices, some RPA tools and protocols may become useful for chatbot QA, but they may also be considered redundant in the presence of more cutting-edge technologies.

6. Discussion

The main limitation of RPA when it comes to chatbot QA and testing is the unpredictable nature of human-chatbot conversations, which is difficult to replicate in a testing environment using RPA-only tools or protocols. AI and ML, augmented by advanced Natural Language Processing (NLP) algorithms that might be capable of discerning intent at a higher degree of accuracy, may prove better at testing most chatbots in the future.

The use of chatbot algorithms for research purposes, which is a different nature of deployment, may be a better use case for RPA-based chatbot QA. We also can't dismiss the possibility of RPA being used for partial testing of chatbots, taking over some aspects of QA and improving the overall process by reducing testing time and human errors.

The best use case for RPA in chatbot QA testing would be for businesses that regularly develop and deploy new chatbots, and many of their RPA-based testing tools can either be directly used to train new chatbots or can easily be repurposed to that end.

The use of RPA in creating and growing testing or training samples is another avenue worth exploring. Even if the RPA cannot directly be applied to end-user testing, it might result in more affordable and easily adoptable solutions for sample-data growth/inflation compared to AI/ML-based tools, especially if the quality of the available sample data is high.

7. Conclusion

Chatbot QA using RPA remains relatively unexplored, partly because of the nature of the testing procedures themselves (especially when it comes to end-users) and partly because AI and ML-based testing tools and QA practices are growing at a much more expedited pace than RPA. They have the potential to make RPA obsolete well before its widespread adoption, at least in the chatbot QA/testing domain. Businesses that are already familiar with and have adopted a wide array of RPA-based tools and practices may find it worth the time and effort to use it for chatbot QA they may be developing or deploying. For businesses that are looking for fresh chatbot QA testing solutions, the alternatives might be better.

References

[1] J. Pereira and O. Diaz, "Chatbot Dimensions that Matter: Lessons from the Trenches," in *Springer*, 2018.

[2] A. Følstad, C. B. Nordheim and C. A. Bjørkli, "What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study," in *Lecture Notes in Computer Science*, 2018.

[3] M. Hermès and V. Juan, *Towards A Better Understanding Of Service Quality Attributes Of A Chatbot,* 2019.

[4] L. Li, K. Y. Lee, E. Emokpae and S.-B. Yang, "What makes you continuously use chatbot services? Evidence from Chinese online travel agencies," *Springer,* 2021.

[5] P. Hofmann, C. Samp, and N. Urbach, "Robotic process automation," *Springer,* 2019.

[6] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications,* vol. 2, 2020.

[7] J. M. López-Morales, P. C. Cañizares, S. Pérez-Soler, E. Guerra and J. d. Lara, "Asymob: a platform for measuring and clustering chatbots," in *ICSE '22 Companion*, Pittsburgh, PA, 2022.

[8] S. Bravo-Santos, E. Guerra and J. d. Lara, "Testing Chatbots with Charm," in *13th International Conference, QUATIC - Springer*, Faro, Portugal, 2020.

[9] J. Jiang and N. Ahuja, "Response Quality in Human-Chatbot Collaborative Systems," in *SIGIR*, Virtual Event, China, 2020.

[10] R. J. Moore and R. Arar, "Conversational UX Design: An Introduction," *Studies in Conversational UX Design,* p. 1 to 16, 2018.

[11] W. M. P. v. d. Aalst, M. Bichler and A. Heinzl, "Robotic Process Automation," *Business & Information Systems Engineering,* vol. 60, p. 269–272, 2018.

[12] L. Ivančić, D. S. Vugec, and V. B. Vukšić, "Robotic Process Automation: Systematic Literature Review," in *International Conference on Business Process Management*, Vienna, Austria, 2019.

[13] R. Syed, S. Suriadi, M. Adams, W. Bandara, S. J. L. C. Ouyang, A. H. t. Hofstede, I. v. d. Weerd, M. T. Wynn and H. A. Reijers, "Robotic Process Automation: Contemporary themes and challenges," *Computers in Industry,* vol. 115, 2020.

[14] K. C. Moffitt, A. M. Rozario, and M. A. Vasarhelyi, "Robotic Process Automation for Auditing," *Journal of Emerging Technologies in Accounting,* pp. 1-10, 2018.

[15] S. Yatskiv, I. Voytyuk, N. Yatskiv, O. Kushnir, Y. Trufanova and V. Panasyuk, "Improved Method of Software Automation Testing Based on the Robotic Process Automation Technology," in *International Conference on Advanced Computer Information Technologies (ACIT)*, Ceske Budejovice, Czech Republic, 2019.

[16] B. Setiaji and F. W. Wibowo, "Chatbot Using A Knowledge in Database," in *7th International Conference on Intelligent Systems, Modelling, and Simulation*, Bangkok, Thailand, 2016.

[17] V. V., J. B. Cooper, and R. L. J., "Algorithm Inspection for Chatbot Performance Evaluation," in *Procedia Computer Science*, Trivandrum, India, 2020.

[18] J. Bozic, O. A. Tazl and F. Wotawa, "Chatbot Testing Using AI Planning," in *IEEE*, Newark, CA, USA, 2019.

[19] M. J. Hloušek, M. Smutek and Z. Hloušková, "SOCIAL COUNSELLING CHATBOT - PILOT

TESTING," in *NORDSCI*, 2021.

[20] A. F. Muhammad, D. Susanto, A. Alimudin, F. Adila, M. H. Assidiqi, and S. Nabhan, "Developing English Conversation Chatbot Using," in *IEEE - International Electronics Symposium (IES)*, Surabaya, Indonesia, 2020.

[21] D. Calvaresi, A. Ibrahim, J.-P. Calbimonte, R. Schegg, E. Fragniere and M. Schumacher, "The Evolution of Chatbots in Tourism: A Systematic Literature Review," in *Springer, Cham - ENTER 2021 eTourism Conference*, 2021.

[22] V. Koumaras, A. Foteas, A. Papaioannou, M. Kapari, C. Sakkas and H. Koumaras, "5G Performance Testing of Mobile Chatbot Applications," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Barcelona, Spain, 2018.

[23] J. Bozic and F. Wotawa, "Security Testing for Chatbots," in *Testing Software and Systems: 30th IFIP WG 6.1 International Conference, ICTSS*, Cádiz, Spain, 2018.

[24] N. Waheed, M. Ikram, S. S. Hashmi, X. He, and P. Nanda, "An Empirical Assessment of Security and Privacy Risks of Web based-Chatbots," in *WISE 2020 Conference*, Amsterdam and Leiden, Netherlands, 2020.

[25] W. E. Hefny, Y. Mansy, M. Abdallah and S. Abdennadher, "Jooka: A Bilingual Chatbot for University Admission," *World Conference on Information Systems and Technologies*, vol. 1367, p. 671–681, 2021.

[26] D. Al-Ghadhban and N. Al-Twairesh, "Nabiha: An Arabic Dialect Chatbot," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.

[27] P. Anki, A. Bustamam and R. A. Buyung, "Comparative Analysis of Performance between Multimodal Implementation of Chatbot Based on News Classification Data Using Categories," *MDPI Electronics*, 2021.

[28] E. Amer, A. Hazem, O. Farouk, A. Louca, Y. Mohamed and M. Ashraf, "A Proposed Chatbot Framework for COVID-19," in *International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo Egypt, 2021.