"Forgot Your Password Again?" - Acceptance and User Experience of a Chatbot for In-Company IT Support

Dario Fiore
Innovation Process Technology AG
Zug, Switzerland
mail@dariofiore.ch

Matthias Baldauf
University of Applied Sciences
St.Gallen, Switzerland
matthias.baldauf@fhsg.ch

Christian Thiel
University of Applied Sciences
St.Gallen, Switzerland
christian.thiel@fhsg.ch

ABSTRACT

Over the last few years, chatbots have become a common channel for customer service interactions. In contrast, their usage for in-company applications and respective scientific knowledge about so-called virtual enterprise assistants from a user perspective are still scarce. In this paper, we studied the acceptance and user experience of a chatbot for in-company IT support. In a user study, 12 employees of a bank and a hospital evaluated and assessed a respective chatbot prototype supporting three typical use cases. Our results indicate that such an in-company chatbot is well-suited for structured use cases, such as resetting a password and releasing an e-mail attachment from quarantine. Participants appreciated the simplicity, the pro-active guidance and immediate feedback. The participants assessed a chatbot and the phone as preferred channels to the IT help desk.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Empirical studies in HCI**.

KEYWORDS

Chatbot, virtual enterprise assistant, user study, IT support, help desk

ACM Reference Format:
Dario Fiore, Matthias Baldauf, and Christian Thiel. 2019. "Forgot Your Password Again?" - Acceptance and User Experience of a Chatbot for In-Company IT Support. In *MUM 2019: 18th International Conference on Mobile and Ubiquitous Multimedia (MUM 2019), November 26–29, 2019, Pisa, Italy*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3365610.3365617

MUM 2019, November 26–29, 2019, Pisa, Italy

ACM ISBN 978-1-4503-7624-2/19/11…$15.00
https://doi.org/10.1145/3365610.3365617

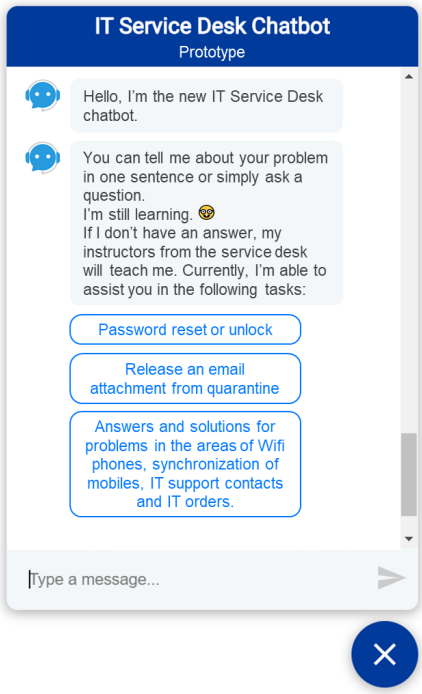

Figure 1: User interface of the chatbot protoype with a welcome message and function selection.

1 INTRODUCTION

"Chatbots are the new apps", Microsoft's CEO Satya Nadella proclaimed back in 2016 [17]. Indeed, over the last few years we have observed the appearance of a large number of respective bots, which enable interaction by human users in natural, written language. The rise is driven by powerful platforms which ease the creation of chatbots, as well as their easy integration into websites and popular messaging apps such *Facebook Messenger*, *Sykpe* or *Telegram* [9]. Overall, the global chatbots market is expected to grow at approximately USD 6 Billion, by 2023 [14].

Currently, we see a major application focus of chatbots on customer service interactions: A lot of e-commerce providers offer chatbots integrated in their Websites, or via messaging platforms, which are able to inform customers about products

and even trigger a purchase in a dialogue-style manner. Since these chatbots are able to initiate transactions, they are often referred to as *virtual customer assistants* (VCA).

Besides these VCA, so-called *virtual enterprise assistants* (VEA) are of increasing relevance, yet are, so far, a less common and studied application area of chatbots: VEA are chatbots for employees for simplifying interaction with in-company systems. In particular, VEAs are supposed to have great potential regarding efficiency increases and cost reductions for in-company IT help desks. Similar to a chatbot for customer interactions, a specialized VEA could assist with common IT-related troubleshooting tasks. While various available chatbots demonstrate the general technical feasibility, scientifically validated knowledge about the employees' attitudes regarding an in-company chatbot is scarce.

In this paper, we investigate the acceptance and user experience of such virtual enterprise assistants offering in-company IT support tasks. We are interested in the suitability of a chatbot for typical IT support cases as well as the experience and potential acceptance of users. To gain respective insights, we conducted a user study involving a sophisticated functional chatbot prototype (Figure 1). The prototype supported three typical IT support tasks (resetting a password, answering frequently asked questions, releasing an e-mail attachment from quarantine). 12 employees of a bank and hospital were confronted with the chatbot, assessed the supported services and compared their experience with traditional help desk channels. The contributions of this research include valuable insights into the overall user experience of an IT troubleshooting chatbot and validated recommendations for suitable use cases.

2 PRIOR WORK

The work presented in this paper builds on two strands of prior research: 1) acceptance and related success factors of chatbots and 2) chatbot applications beyond traditional e-commerce scenarios.

Acceptance and Success Factors of Chatbots

What makes a chatbot accepted by users in general and what are the respective success factors has been subject of research in several recent publications (cf. [4, 5, 7, 15, 16, 22, 25]). In the following, we provide a brief overview of respective study results and corresponding recommendations from prior work.

A fundamental requirement for the establishment of trust in a chatbot and thus its acceptance by users, is the *transparency of a chatbot* [4, 7, 15]. A chatbot must clearly indicate that its user is not chatting with a human being, but a robot. Hence, a chatbot should introduce itself at the beginning of a conversation and point out its nature.

During the conversation, misunderstanding of the chatbot's functionality can lead users to posing non-answerable questions and to frustration, as a consequence. Thus, a chatbot should *clearly communicate its features* in order to correctly set the users' expectations [4, 5, 7].

Further recommendations include a *proactive conversation style* of a chatbot [7, 25]. This means that a chatbot should actively guide the conversation to support the user. A rather passive chatbot which mainly waits for user input, runs the risk of losing context within conversation. An efficient chatbot should help the user with concrete questions and, as often as possible, provide predefined answer opportunities. These can be presented as buttons which increase the efficiency and reduce misunderstandings [3, 7]. Furthermore, buttons, animations and videos that are carefully integrated in a chatbot are perceived as natural and engaging [7, 16].

The language of the chatbot should be *suitable for the target group* and fit the requirements of its users. Additionaly it is recommended to avoid overly long and complex sentences and split up longer statements of a chatbot into several shorter sentences [15, 16, 25]. In a similar vein, the personality of a chatbot should be consistent and fit the corresponding use case and context [5, 7, 16, 22, 25].

The key motivator for using a chatbot turned out to be the *expected increased productivity* [3]. Users hope to receive support and information easily and quickly. Thus, the effectiveness and efficiency of a chatbot and its services need to be ensured.

Highly relevant for the IT support application is the recommendation to *provide escalation opportunities to a human expert*. In case a chatbot cannot understand and deal with the user's request, the chatbot should either offer to handover or directly forward to a human specialist [15, 16, 22, 25].

Chatbot Applications beyond E-Commerce

Several companies already make use of chatbots to provide information about their services and products and let customers initiate purchases. Examples include *Uber*, *Pizza Hut*, *Marriott* and the airline *SAS*, which offer requesting a ride, placing orders for pizza and booking hotel rooms and flights, respectively, within the *Facebook Messenger*.

While such traditional e-commerce services are publicly available, recent research aims at advancing the functionality of chatbots and exploring novel application domains for chatbots. One relevant domain is banking with complex consulting services and special privacy and security demands. For example, Lai et al. [10] introduce their "chatbot security control procedures" for improving the security of a banking chatbot and reduce potential security risks. Letheren and Dootson[12] draw comparisons to online banking, where customers are prepared to trade security for convenience. They conclude that bank chatbots need to be designed to give a "sense of security".

Another recent usage domain for chatbots beyond typical virtual customer assistants in e-commerce are e-government services. Androutsopoulou et al. [2] present chatbots as a new digital channel of communication between citizens and government for information provision, for example. They argue that, in comparison to existing digital communication channels, chatbots enable a richer and more expressive interaction of citizens with government in everyday language, facilitating and advancing both information seeking and conducting of transactions. Porecca et al. [18] propose to employ chatbots for facilitating access to open governmental data. They present a respective prototype and argue that chatbots might provide an intuitive interface to such data when no specific user-friendly applications are built on top of them.

Scientific studies on in-company chatbots are still scarce. One of the rare examples is presented by Liao et al. [13] who studied a field deployment of a "Human Resource" chatbot providing company-related information assistance to new employees. The researchers focused on the investigation of the users' conversational interactions, and did not report on the overall acceptance and user experience of the prototype. Toxtli et al. [24] presented the in-company chatbot "TaskBot" for improving collaboration. They investigated how such a chatbot can help mediating task manangement for individuals and teams. Users could simply mention the chatbot in a conversation and delegate tracking of tasks to it. According to the results, the participants enjoyed the simplicity of creating and tracking tasks from the conversation and especially appreciated that no further tool for tracking was required.

Prior research on digital assistants supporting natural language for troubleshooting focused on spoken language and technical aspects. For example, Acomb et al. [1] presented an early state-of-the-art report on technical support systems using spoken dialogues. A more recent overview of chatbots for troubleshooting is provided by Thorne [23] who outlines key properties of troubleshooting dialogues, respective technical architectures and evaluation strategies. Scientifically validated knowledge on in-company chatbots for IT troubleshooting from an employee perspective is scarce.

3 EXPERIMENT

To learn about the acceptance and user experience of employees with an in-company chatbot, we conducted a user study involving an advanced functional chatbot prototype.

Prototype

We implemented our chatbot prototype using *Rasa Stack*[1], an open source conversational AI platform. Rasa is written in Python and applies advanced machine learning for text recognition and dialog management [21]. The prototype makes use of three core components: *Rasa-NLU* for recognizing text entry (with the classification library *sklearn* and a pretrained language model using *spaCy*), *action server* for development and execution of custom actions and form actions, and *Rasa-Core* for dialogue management and central node for various further components. For alternative implementation techniques, see the technical overview in [19]; general information on how to implement a chatbot can be found in [8], e.g.

The chatbot user interface was integrated in a Web page template and thus, was accessible through a standard Web browser. When the chatbot was started, it welcomed the user and introduced its purpose in brief (Figure 1). It offered three use cases as buttons to ease selecting one. When a use case was completed or cancelled, the chatbot asked to provide further assistance and offered the list of use cases again. If the user decides to quit (by typing "bye" or similar), the chatbot said good bye with a random farewell message. When the user required help and typed a respective query (or used the help button displayed at particular times), the chatbot describes its functions again (without newly introducing itself). Overall, throughout the design of the chatbot, the previously described success factors (see Section 2) were considered. For example, the chatbot transparently states that it is a robot, not a human expert, it pursues a proactive conversation style, etc.

Tasks

We aimed for a high ecological validity of the study and thus carefully selected realistic application scenarios for the chatbot. We targeted very frequently used support tasks of IT service desks (i.e., participants are probably familiar with these use cases and have been in a respective scenario themselves before). The tasks were identified from interviews with the heads of two company IT service desks, from talks with experienced representatives of the chatbot platform *Rasa* and the chatbot consulting agency *Materna*, as well as from respective available literature [15, 20, 25].

Based on these sources, we selected three frequently mentioned use cases for implementation: *resetting password* (mentioned by all sources), *answering FAQs* (mentioned by the service desk heads and Rasa), and *releasing e-mail attachments* (mentioned by the service desk heads).

We describe the implementation of these use case in more detail in the following. It is noteworthy, that each of them assumes that the chatbot knows the current user (i.e., the employee is logged in, typically through a "single sign-on" solution in the intranet).

Resetting Password. The reset of the password of a business user account is very common task of an IT service desk, when a user either cannot recall it or the account was locked

[1] http://rasa.com

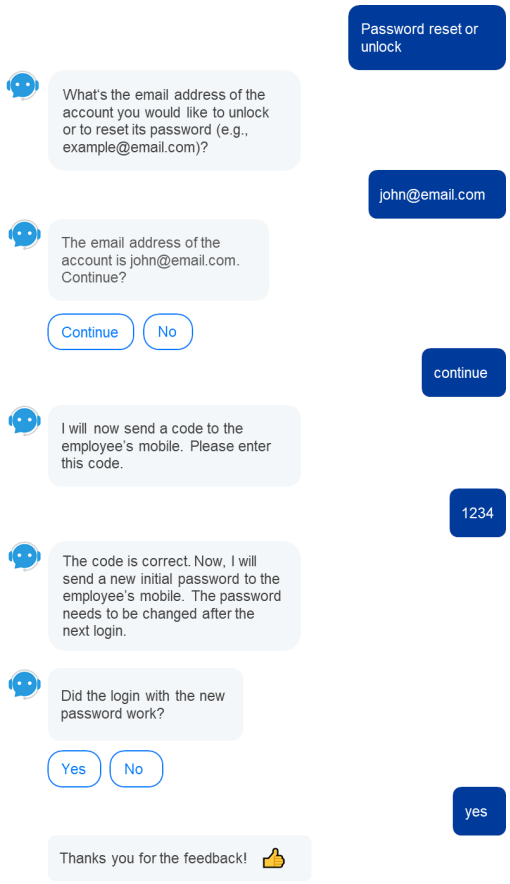

Figure 2: Successfully resetting a password.

(typically due to too many incorrect login attempts). Enabling employees to reset their password independently is a crucial feature to ensure work continuation also outside support hours. Since this use case is obviously security-critical and its misuse has to be prevented, the following real-life requirements were defined: One employee alone may not initiate the password reset, two-factor authentication needs to be applied, and the new password must not appear in the chatbot protocol.

The (simplified) typical workflow of this use case is as follows (depicted in Figure 2):

(1) The employee (E1), who is not able to login, asks a colleague (E2) to initiate the password reset.
(2) E2 starts the chatbot (and is authenticated through Single-Sign-On) and selects the "password reset" feature.
(3) The chatbot asks E2 for the e-mail address of E1.
(4) A text message with a code (second factor) is sent to the mobile phone of E1.
(5) The chatbot asks E2 for the code.
(6) E2 enters the correct code received from E1.
(7) A text message with a temporary password is sent to the mobile phone of E1 and the E1's account is unlocked (the chatbot informs about these steps).
(8) The chatbot asks whether logging-in with the new temporary password was successful. If not, it refers to the IT service desk. If yes, it asks for any further requests (see above).

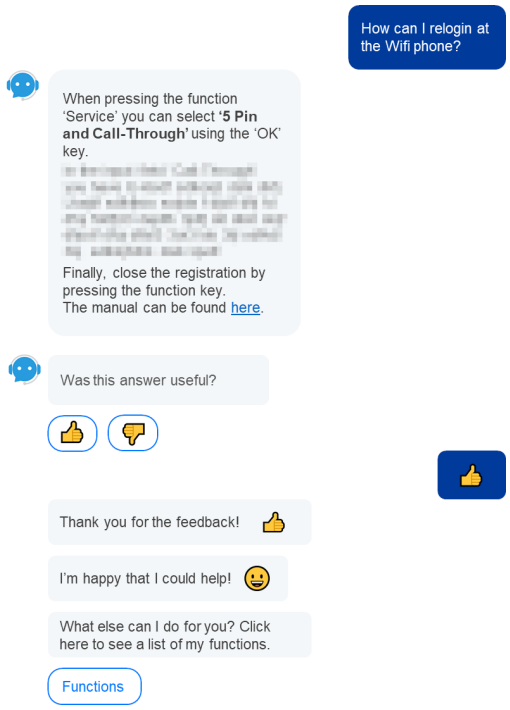

Figure 3: Successfully answering a frequently asked question.

Answering FAQs. According to the involved service desk heads, many requests can be simply answered by standardized replies. This use case should therefore recognize frequently asked questions and provide suitable, predefined answers. In order to guide the users' expectations and provide a realistic setting, we defined the following requirements: The chatbot should communicate the topics it currently can provide support with and, to allow for a continuous improvement, it should ask the user, whether the answer was suitable and helpful or not.

The (simplified) typical workflow of this use case is as follows (depicted in Figure 3):

(1) The employee types a question.
(2) The chatbot provides an answer.
(3) The chatbot asks whether the answer was suitable and helpful.
(4) The employee types "yes" or "no" (or pushes the respective button).

(5) The chatbot provides an appropriate reply: If the chatbot's answer was considered helpful, it replies "I'm happy that I could help :-)", otherwise it apologizes and points out that this prototype is continuously improved.

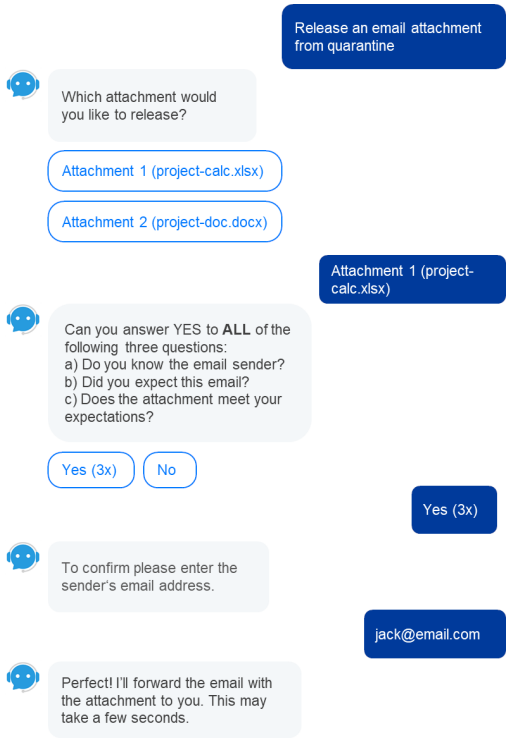

Figure 4: Successfully releasing a e-mail attachment from quarantine.

Releasing E-mail Attachments. Like in many typical company settings, file attachments of incoming e-mails are scanned for malicious content at the involved companies. In the case of such being found, specialized filters recognize potentially harmful attachments, respective attachments are not delivered to the recipients but blocked and need to be manually released by the IT support on request of the recipient. Typical questions for assessing the risk include whether the recipient knows the sender, expected the message, and considers the message content reasonable. Only if an employee answers yes to all these questions, should the respective attachment be released.

The (simplified) typical workflow of this use case is as follows (depicted in Figure 4):

(1) The employee types that he would like to release an e-mail attachment (or selects the corresponding button).
(2) The chatbot presents a list of file attachments (as buttons with file name including file extension), which are currently in quarantine for this user, and asks which one should be released.
(3) The employee presses the respective button.
(4) The chatbot poses three relevant questions to raise awareness regrading a potential security threat.
(5) The employee types "yes" or "no" (or pushes the respective button).
(6) If the employee selected "no", the chatbot asks to contact the IT service desk. If he selected "yes" , the chatbot asks to enter the sender's address of the respective e-mail as a final security check.
(7) The employee enters the sender's e-mail address.
(8) The chatbot provides an appropriate reply: If the e-mail address is correct, the attachment is released and forwarded to the employee. If it is not correct, it offers two more opportunities to enter the correct e-mail address, then cancels this use case with a respective message to the employee.

Table 1: Participants Characteristics

Id	Comp.	Role	Sex	Age
P1	Bank	HR Developer	f	21
P2	Bank	Software Engineer	m	43
P3	Bank	Dep. Head IT Service Desk	m	40
P4	Bank	Private Client Advisor	f	31
P5	Bank	Team Lead Services	f	31
P6	Bank	Private Client Advisor	m	58
P7	Hospital	Trainee Corp. Comm.	m	26
P8	Hospital	Specialist IT Training	f	41
P9	Hospital	Specialist IT User Mgmt.	f	52
P10	Hospital	Assistant	f	24
P11	Hospital	Assistant	f	25
P12	Hospital	Assistant Doctor	m	24

Participants

In order to gain valid and practically relevant feedback, we conducted the user study not with an artificial sample, but with representatives of the prospective user group. We recruited volunteer employees from a bank and a hospital, since these employees are used to dealing with security- and privacy-relevant information. During recruitment, we aimed at a heterogeneous group of study participants (with regard to sex, age, and job position).

Two crucial requirements for the participants were language proficiency as well as computer literacy. We only selected volunteers who spoke German at level C2 ("Mastery or proficiency") and were familiar with operating desktop computers, in order to preclude linguistic misunderstandings,

and hurdles and troubles regarding keyboard and mouse input.

Overall, 12 participants took part in our study (7 female, 5 male). They were aged between 21 and 58 (M=34.7; SD=11.5). Table 1 shows the participants characteristics. On a five-point Likert scale (1-"I am not experienced in using computers at all", 5-"I am very experienced in using computers"), 7 participants rated their IT experience with 5; 4 participants with 4; and one participant with 3. 2 participants had not used a chatbot before. 9 participants had used chatbots one to 5 times, one person more than 5 times. All of the participants were familiar with the currently offered help desk channels in their company and had contacted the help desk before, at least once. While 3 of the involved bank employees use *Skype for Business*, there was no related chat/communication application available for the hospital employees.

Study Design and Setup

Our study was designed as a within-subject experiment, i.e. each participant tested and assessed each of the three chatbot use cases. Up to three participants took part in a test session in parallel. For each one, a separate notebook computer was provided. These identical computers ran the described chatbot prototype and were used to presenting and filling in the involved questionnaires. The study assistant was present throughout the entire sessions and took care that the participants did not talk to each other during the execution of the study tasks and the filling in of the questionnaires.

For improved ecological validity, the test sessions were conducted in meeting rooms at the involved companies by a study assistant. Each session took between 60 and 90 minutes.

Introduction. Each test started with a brief introduction by the study assistant. He welcomed the participants and explained the pursued research goals and the study procedure. By means of a short questionnaire, we wanted the participants to state their current knowledge ("I know what a chatbot is and how to use one." - on a five point Likert scale from 1-no to 5-yes) and experience ("How often did you use a chatbot before?") regarding chatbots.

Then, the study assistant asked the participants to sign the study consent allowing us to evaluate and publish the data to be collected for scientific purposes.

Afterwards, the study assistant introduced an available e-commerce chatbot[2] and encouraged the participants to experiment with it for about five minutes. During this time, the study assistant answered arising questions.

Chatbot Usage. The study assistant kicked off the main phase of the study by distributing a fictional news article on the company-internal news page. The article announced the new "IT service desk chatbot" and briefly introduced its three main use cases (see "Tasks").

The study assistant then asked the participants to start the chatbot and solve the given tasks (resetting password, answering FAQs, releasing e-mail attachments from quarantine). Each participant was provided with a separate task order, i.e. none of the participants in one test session solved the same task at the same time. The order of the tasks was systematically varied using a balanced latin square in order to counterbalance and avoid learning effects.

When the participants had tested one chatbot use case for about 10 minutes, they were asked to complete a short questionnaire regarding this specific use case. It included the statements "I would use the chatbot to solve this task" and "Using the chatbot I could solve this task efficiently", both with a five-point Likert scale (1-no; 5-yes). Furthermore, we asked for positive and negative remarks and potential for improvements.

Final Questionnaire. Having experienced and tested each of the three use cases, the participants were asked to complete a final questionnaire.In this questionnaire, they could state additional use cases they considered suitable for such an IT troubleshooting chatbot. Furthermore, they were asked to rate their overall willingness (independent of a concrete use case) to use a respective chatbot for IT support and to rank different communication channels to the company IT help desk regarding their personal preference (chatbot, ticket system, phone, etc.). Additionally, we wanted to know over which channels the participants would like to use such a chatbot (web-based chat integrated in an intranet portal, as Skype integration, etc.).

Finally, we asked the participants to assess their overall user experience on the *User Experience Questionnaire* (UEQ) by Laugwitz et al. [11]. This common questionnaire allows the assessment of an application according to the six dimensions Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Furthermore, the corresponding online tool allows a direct comparison with previously assessed applications.

The survey closed with questions regarding demographic data (age, sex, job position) and a self-assessment of the own IT literacy.

4 RESULTS

This section presents the results of our user study. In the first three subsections, we report on user feedback for the different use cases of the prototype. Then, we summarize the results of a comparative assessment and report on the overall acceptance and generated ideas for further use cases.

[2]https://www.otto.de/clara/

Task 1: Resetting Password

Positive Remarks. 7 of the 12 participants expressed their astonishment regarding the speed of the chatbot, 3 positively mentioned its simplicity.

P1 appreciated that *"the chatbot apologizes and recommends contacting the service desk, if the problem is not solvable"*. P8 liked the two-factor authentication via text message which *"prevents resetting the password and thus the stealing of data by a stranger"*. P10 stated positively, that *"the chatbot helps to solve the task step by step"*.

Negative Remarks. The participants had almost no negative remarks regarding this use case. One participant noted that *"It would be easier if I knew more alternative expressions for this problem."* (P3)

Task 2: Answering FAQs

Positive Remarks. 3 of the 12 participants explicitly mentioned the pleasant speed of the chatbot for solving this task. For example, *"the chatbot allows for solving this task quickly and straightforwardly"* (P1)

P1 and P3 appreciated that *"images and screenshots were shown"*.

P2 pointed out another central aspect of a chatbot in comparison with a human representative of the IT help desk: *"I don't annoy anybody but still receive support"* (P2)

P7 emphasized the integration of the available ticketing system into the chatbot: *"Awesome, especially the provided links to our ticket system and to manuals - this makes such things much easier."*

Negative Remarks. 3 participants mentioned that a few keywords should be enough to "ask" the chatbot, instead of having to phrase entire questions.

P10 would like to have support to recognize whether the chatbot's understanding was correct: *"I would like to see in the answers whether the chatbot understood me correctly, for example, by repeating parts of my question."*

P1 and P3 criticized the chatbot's user interface in the case of long responses, as indicated by the following statement: *"Some answers contain much text in this small speech bubble. This requires scrolling many times, which is rather cumbersome."* (P1)

When conducting task 2, P12 had a note regarding the potential overall acceptance in the hospital domain: *"I'm used to looking up a phone number in the intranet and calling the person. If the number is wrong, I usually get forwarded correctly. In the medical domain, there's a low threshold for using the phone, so introducing a chatbot in this area might be rather difficult."*

6 participants criticized that their question was not understood correctly in the first try. P7 mentioned that he would like to receive pro-active help from a human representative of the help desk: *"In such a case, a specialist should call the employee or start a chat."* (P7)

Task 3: Releasing e-mail Attachments

Positive Remarks. Again, 5 of the 12 participants emphasized the speed of the chatbot for this use case: *"This is much faster than writing a ticket."* (P2) or *"Works very fast. These questions can be answered via computer very well."* (P4)

2 participants were explicitly positive about the step by step instructions of the chatbot. P6 mentioned that it is *"simple to use and easy to understand."*

Negative Remarks. The rare negative remarks during conducting this task focused on the actual navigation of the chatbot rather than on the specific task itself. For example, P1 remarked: *"It wasn't clear for me how to start this function of the chatbot. The option menu should be visible all the time."*

Comparative Assessments

Figure 5 shows the participants' answers regarding the statement "I would use a chatbot to solve this task". The participants rated the use case "resetting password" very positively: all of them agreed with "rather yes" or "yes". "Answering FAQ" was perceived differently by the participants: 6 stated that they would rather not use a chatbot for this task, 2 were indecisive, 4 stated "rather yes" or "yes". Finally, 10 participants were positive about releasing an e-mail attachment via the chatbot ("rather yes" or "yes") while two participants would (rather) not use a chatbot for this task.

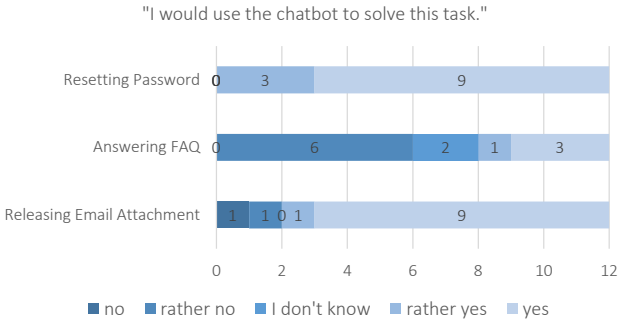

Figure 5: The participants' willingness to use a chatbot at work in general.

Overall Acceptance and Further Use Cases

We also asked the participants about their overall willingness to use a chatbot at work (i.e., not specifically for IT support tasks). As depicted in Figure 6, only two participants were undecided, the remainder (83%) were (rather) positive about a chatbot for professional usage: 6 participants stated "yes", 4

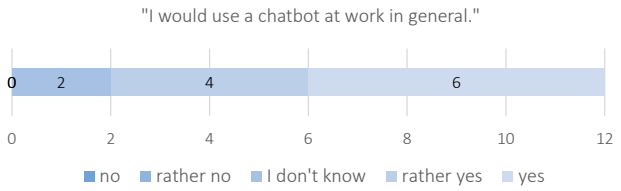

Figure 6: The participants' willingness to use the evaluted chatbot for the three tasks.

participants "rather yes". For example, supportive statements included *"the chatbot enables quick answers, long searches could be shortened"* (P5) and *"this would be very useful and helpful"* (P11).

Beyond the use cases covered by the three test tasks, our participants imagined several additional useful use cases for an in-company chatbot. Additional services that were mentioned for the IT service desk include:

- Managing access rights after having lost the staff badge
- Restoring a file
- Step-by-Step guidance for several IT tasks for beginners
- Overall technical support
- Reporting defects

Furthermore, the participants came up with several useful services of an in-company chatbot beyond IT service tasks, such as:

- Looking up phone numbers of colleagues
- Looking up rooms in a large company buildings
- Information on various opening times of offices
- Looking up time schedules (e.g. surgery in the hospital)
- Answering questions regarding legal regulations
- Ordering office material or documents
- Changing workplans and task schedules

For many of these tasks, the corresponding information (phone numbers, opening times, etc.) is available on the intranet portals of the bank and the hospital. However, the participants considered such look-up tasks as convenient applications for a chatbot.

Preference of Communication Channels

To investigate the participants' preferred ways to communicate with the in-company IT service desk, we asked them to order six suitable communication channels (1 - least preference; 6 - highest preference). We summed up the assessments per channel and normalized the sums within a range from 0 to 1. The results are depicted in Figure 8: Overall, the chatbot solution was rated best with a normalized rating of 0.71, followed by phone (0.64). Communication via e-mail, self service using manuals and tutorials from the intranet, and personal communication were rated similarly with values of 0.54 and 0.55. Finally, a typical ticket solution for managing incidents was rated worst (0.45).

Table 2: Selected items of the UEQ with remarkable ratings

Item	Mean	Var.
easy / hard to learn	1.9	0.6
valuable / inferior	1.9	2.0
fast / slow	2.7	0.4
obstructive / supportive	2.1	1.2
organized / cluttered	2.1	0.8
conservative / innovative	2.1	0.8
unpredictable / predictable	0.5	1.4

Regarding the integration or provisioning of a chatbot (with)in the company, we asked the participants to assess three alternatives. 9 participants stated that they would like to use such a chatbot as a webchat integrated into the available intranet portal. An integration into *Skype for Business* was liked by 7 participants, the solution of a custom chat app for smartphones by 5 participants. For each of the latter two options, 3 participants explicitly refused a respective offer.

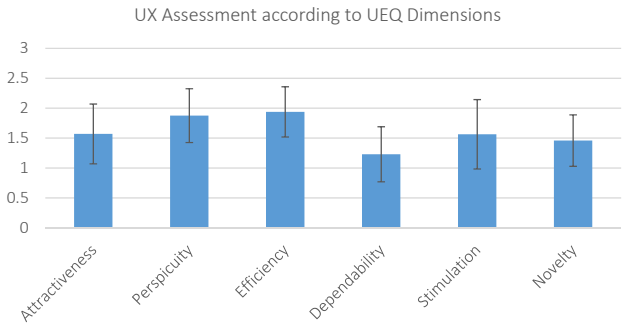

Figure 7: Results of the UX evaluation according to the six dimensions of the User Experience Questionnaire (UEQ).

User Experience Assessment

Figure 7 shows the results for the six UEQ dimensions. None of the dimensions is rated negatively (-3 to 0), but the rating of each dimension is in the positive range (with values between 1.23 and 1.94).

In comparison to mean values of prior 401 evaluations with the UEQ, the chatbot prototype performed very well. In five dimensions the user experience of the chatbot was rated above average, only its *Dependability* was at average.

Table 2 shows selected items of the UEQ with remarkable results. The item *fast/slow* received the highest mean of 2.7 (very low variance of 0.4), i.e. the chatbot was perceived as

very fast. Furthermore, with a high means of 2.1, the chatbot was perceived as *supportive*, *organized*, and *innovative*, with a means of 1.9 as *easy to learn* and *valuable* (while these ratings showed a high variance of 2.0). The item *unpredictable/predictable* was the only item rated with a mean below 0.8, i.e. representing a neutral evaluation.

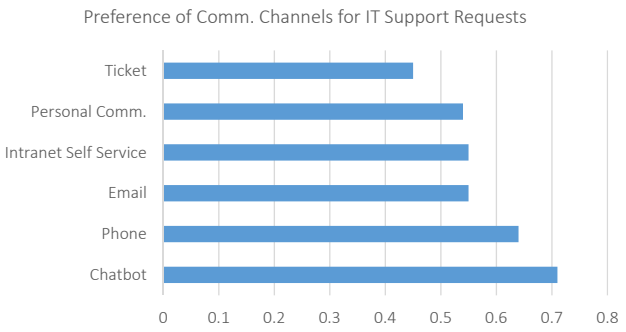

Figure 8: The participants' preferred channels for communicating with the IT service desk.

5 DISCUSSION

We structured the discussion of our results along three main objectives of our investigation: the overall chatbot user experience, the suitability of use cases as well as the comparison with other help desk channels and a respective integration.

Overall Chatbot User Experience

Overall, the participants were positive about the evaluated chatbot prototype. They appreciated that it is easy to learn and fast. Speed, and thus efficiency, was one of the main advantages mentioned by the participants. Several participants referred back to prior experiences with the traditional in-company IT support (writing an e-mail, submitting a support ticket, waiting for a return call, e.g.), and emphasized the advantage of immediate synchronous feedback by such a chatbot solution.

The main source of negative emotions of the participants were misclassifications of their input. Several participants expressed their frustration during task 2 (answering FAQs), where they had to rephrase their original request since the chatbot did not understand their first entry.

This drawback also became evident in the quantitative UX evaluation: While its results confirmed important aspects for the investigated domain such as *easy to learn*, *fast*, *organized*, and *supportive*, the participants were indecisive about its *predictability*. Analyzing the participants' remarks during the study, we clearly ascribe this outcome to the aforementioned misunderstandings, particularly in task 2.

Suitability of the Use Cases

Two use cases, resetting a password and releasing an e-mail attachment, received highly positive feedback with almost no negative remarks. As mentioned above, remarks on the second use case, answering FAQs, were more diverse. While several participants also appreciated the speed for solving this task, half of the participants mentioned that they had to rephrase their question to receive a correct answer. Additionally, the participants appreciated supporting images and links in this use case, yet they disliked overly long text and images, which required scrolling.

The direct comparative assessment of the use cases ("I would use the chatbot to solve this task.") confirmed the participants' preference for *resetting password* and *releasing an e-mail attachment*. This result is in line with related studies on chatbots in other domains which conclude proactive guidance ([7], e.g.) and recommend well-structured use cases for chatbots. Prior research (e.g. [9]) also reports on simple misunderstandings as one of the main reasons for failing dialogues, despite recent advancements in the field of natural language understanding. In our study, we observed the great variability in phrasing questions especially in task 2, even when the "chatbot's knowledge" is explicitly mentioned (such as "Wifi phones, synchronization of mobiles, ..." in Figure 1). For *answering FAQs* (and related use cases including open questions) a very high quality of natural language understanding is required to provide helpful feedback (particularly, if potential questions are narrow and similar). We consider this especially relevant for the investigated scenario of IT support, where, due to preceding issues with IT systems, users might already be frustrated or upset when using a respective chatbot. Misunderstandings and incorrect answers could further increase their negative emotions and, as a further consequence, impact their job performance.

To avoid such scenarios and provide a robust real-life implementation, more training data is required to improve the chatbot's understanding and mechanisms should be integrated to allow for continuous learning and improvement over time. Furthermore, a chatbot should ask the user for clarification with a few most likely options if it cannot determine the intent of the user with high probability. Our research prototype was not used on a large scale and thus did not include such features to keep implementation efforts reasonable.

Alternatively, chatbot use cases with open questions could benefit from recent suggestions such as a "Context View" [6] visualizing the chatbot's state of understanding. Furthermore, we learned that the chatbot answers in this trouble shooting domain, due to their complexity for some requests, need to be carefully designed. Participants appreciated links to existing systems and manuals and disliked longer chatbot statements

which required scrolling. Preparing comprehensive technical instructions in the form of compact chatbot answers seems challenging (and costly), thus supported services should be carefully selected.

As a first use case for realizing and introducing a respective in-company chatbot, resetting a password seems a good first use case. While its usage was assessed very positively by our participants, it, furthermore, promises a significant relief for the IT service desk: Gartner estimates that requests for resetting a password account for 40% of the overall IT service desk requests [26].

Comparison to Alternative Channels and Integration

When being directly compared to alternative channels to the IT service desk, the chatbot approach was well-received by our participants. In contrast to common channels such as a ticketing system, e-mail and phone support, a chatbot solution ensures an immediate and quick response - which was highly appreciated by our users. While a chatbot may not replace the traditional IT help desk today (especially for complex open questions, see above), it might provide a completion for existing offers. Well-structured use cases for first level support requests may relieve the IT support team from a significant work load.

Regarding the integration of the chatbot, we saw the preference tending to an integration into the companies' intranet portals. Typically, employees are very familiar with the intranet portal since they use it regularly to read internal news, download forms, or record their working hours. Furthermore, such an integration on a Web page is increasingly well known from related customer service chatbots.

Dependent on the availability and popularity of an in-company messaging tool, consideration could be given to integrating the troubleshooting chatbot into this tool. In our study, the participants employed at the bank, were familiar with *Skype for Business* and approved a respective integration. The hospital employees did not use a company-wide (real-time) messaging tool and thus had problems imagining using use the chatbot over such a channel.

Finally, our participants saw only limited benefits of a smartphone app for the troubleshooting chatbot. Their remarks suggest that there are only rare cases when accessing such a support chatbot on a mobile device would be needed.

Limitations

The involved participants were carefully selected from a pool of volunteer employees of the two involved companies. Since we put emphasis on recruiting real employees (familiar with existing help desk channels as well as special enterprise requirements regarding privacy and security, e.g.), the study features a limited sample size of 12 participants. We thus do not claim statistical proof of the study results, yet aim for identifying trends and tendencies.

The majority of our participants felt confident using a computer. While we see these participants as suitable representatives of the target group, we would like to note that results of a less computer-savvy participant group might differ from the presented ones. In a large company there are typically groups of employees with diverse backgrounds, thus respective results should be generalized with caution for an entire company.

As mentioned in the discussion, there are several ways to evolve the prototype towards a productive version and improve its robustness. Obviously, more training data would further enhance the quality of the chatbot answers.

6 CONCLUSION AND OUTLOOK

In this work, we investigated the acceptance and user experience of a chatbot for in-company IT support. We conducted a user study involving 12 employees of a bank and a hospital who evaluated and assessed a respective troubleshooting chatbot. The sophisticated prototype was based on a state-of-the-art chatbot engine and utilized real-world training data provided by two IT help desks. Its dialogues and language were designed according to results of prior related research.

Our results indicate that such an in-company chatbot is well-perceived by employees. In particular, structured use cases such as resetting a password and releasing an e-mail attachment from quarantine, where the chatbot actively guides the user through the process, received very positive feedback. Participants appreciated the simplicity, the pro-active guidance and immediate feedback. Besides positive feedback regarding its speed, the use case of answering frequently asked questions also aroused negative remarks regarding misunderstandings. The recognition of such open questions is more error-prone and may easily lead to frustration. Still, our participants assessed a respective chatbot and the phone as preferred channels to the IT help desk.

The highly structured use cases, which were favored by our participants, could also be implemented in the form of step-by-step "wizards" in an internal self-service portal. Future research should investigate how a chatbot would perform against such a more formal self-service offer and study the impact of the dialogue-style communication, e.g. Furthermore, we consider long-term studies of similar deployments in the field promising, in order to assess the impact of novelty effects and unveil combined strategies involving alternative channels to the IT help desk.

ACKNOWLEDGMENTS

We would like to thank the heads of the two involved help desks as well as the study participants for contributing their valuable time.

REFERENCES

[1] Kate Acomb, Jonathan Bloom, Krishna Dayanidhi, Phillip Hunter, Peter Krogh, Esther Levin, and Roberto Pieraccini. 2007. Technical Support Dialog Systems: Issues, Problems, and Solutions. In *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies (NAACL-HLT-Dialog '07)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 25–31. http://dl.acm.org/citation.cfm?id=1556328.1556332

[2] Aggeliki Androutsopoulou, Nikos Karacapilidis, Euripidis Loukis, and Yannis Charalabidis. 2019. Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly* 36, 2 (2019), 358 – 367. https://doi.org/10.1016/j.giq.2018.10.001

[3] Petter Bae Brandtzaeg and Asbjørn Følstad. 2017. Why people use chatbots. In *Internet Science*.

[4] Petter Bae Brandtzaeg and Asbjørn Følstad. 2018. Chatbots: Changing User Needs and Motivations. *Interactions* 25, 5 (Aug. 2018), 38–43. https://doi.org/10.1145/3236669

[5] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *INSCI*.

[6] Mohit Jain, Ramachandra Kota, Pratyush Kumar, and Shwetak N. Patel. 2018. Convey: Exploring the Use of a Context View for Chatbots. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 468, 6 pages. https://doi.org/10.1145/3173574.3174042

[7] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 895–906. https://doi.org/10.1145/3196709.3196735

[8] Rashid Khan and Anik Das. 2017. *Build Better Chatbots: A Complete Guide to Getting Started with Chatbots* (1st ed.). Apress, Berkely, CA, USA.

[9] Lorenz Cuno Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 555–565. https://doi.org/10.1145/3064663.3064672

[10] Sen-Tarng Lai, Fang-Yie Leu, and Jeng-Wei Lin. 2019. *A Banking Chatbot Security Control Procedure for Protecting User Data Security and Privacy*. 561–571. https://doi.org/10.1007/978-3-030-02613-4_50

[11] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[12] Kate Letheren and Paula Dootson. 2017. Banking with a chatbot: A Battle between convenience and security. https://eprints.qut.edu.au/114516/

[13] Q. Vera Liao, Muhammed Mas-ud Hussain, Praveen Chandar, Matthew Davis, Yasaman Khazaeni, Marco Patricio Crasso, Dakuo Wang, Michael Muller, N. Sadat Shami, and Werner Geyer. 2018. All Work and No Play?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 3, 13 pages. https://doi.org/10.1145/3173574.3173577

[14] Market Research Future. 2019. Chatbots Market Research Report-Global Forecast 2023. https://www.marketresearchfuture.com/reports/chatbots-market-2981 Accessed 2019-07-06.

[15] Chris Matchett, Rich Doheny, Magnus Revang, and Kenneth Gonzalez. 2017. When Will AI Virtual Support Agents Replace Your IT Service Desk? https://www.gartner.com/document/3689417?ref=ddrec&refval=3865567 Accessed 2019-05-27.

[16] Frédéric Monard, Hans-Peter Uebersax, Jaouad Mousser, Pia Furchheim, Steffen Müller, and Frank Hannich. 2018. Chatbot-Studie. Die digitalen Helfer im Praxistest. https://www.pidas.com/chatbot-studie Accessed 2019-05-27.

[17] Blake Morgan. 2017. How Chatbots Will Transform Customer Experience: An Infographic. https://www.forbes.com/sites/blakemorgan/2017/03/21/how-chatbots-will-transform-customer-experience-an-infographic/

[18] Simone Porreca, Francesco Leotta, Massimo Mecella, Stavros Vassos, and Tiziana Catarci. 2018. Accessing Government Open Data Through Chatbots. In *Current Trends in Web Engineering*, Irene Garrigós and Manuel Wimmer (Eds.). Springer International Publishing, Cham, 156–165.

[19] Kiran Ramesh, Surya Ravishankaran, Abhishek Joshi, and K. Chandrasekaran. 2017. A Survey of Design Techniques for Conversational Agents. In *Information, Communication and Computing Technology*, Saroj Kaushik, Daya Gupta, Latika Kharb, and Deepak Chahal (Eds.). Springer Singapore, Singapore, 336–350.

[20] Magnus Revang and Van Baker. 2018. Four Use Cases for Chatbots in the Enterprise Now. https://www.gartner.com/document/3879492?ref=ggrec&refval=3879492 Accessed 2019-05-27.

[21] Magnus Revang, Van Baker, Brian Manusama, and Anthony Mullen. 2018. Market Guide for Conversational Platforms. https://www.gartner.com/document/3879492?ref=ggrec&refval=3889098 Accessed 2019-05-27.

[22] Toni Stucki, Sara D'Onofrio, and Edy Portmann. 2018. Chatbot – Der digitale Helfer im Unternehmen: Praxisbeispiele der Schweizerischen Post. *HMD Praxis der Wirtschaftsinformatik* (04 2018). https://doi.org/10.1365/s40702-018-0424-8

[23] Camilo Thorne. 2017. Chatbots for troubleshooting: A survey. *Language and Linguistics Compass* 11 (10 2017), e12253. https://doi.org/10.1111/lnc3.12253

[24] Carlos Toxtli, Andrés Monroy-Hernández, and Justin Cranshaw. 2018. Understanding Chatbot-mediated Task Management. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 58, 6 pages. https://doi.org/10.1145/3173574.3173632

[25] Matt Velloso, Kamran Iqbal, Cathy Qian, and Robert Standefer. 2018. Create task automation bots - Bot Service. https://docs.microsoft.com/en-us/azure/bot-ser-vice/bot-service-design-pattern-task-automation Accessed 2019-05-27.

[26] Neil Wynne and Brian Iverson. 2014. Market Guide for Password Management Tools. https://www.gartner.com/en/documents/2761018 Accessed 2019-05-27.