



# **“I’m Disabled and Married to a Foreign Single Mother”. Public Service Chatbot’s Advice on Citizens’ Complex Lives**

Linett Simonsen, Tina Steinstø, Guri Verne, Tone Bratteteig

## **► To cite this version:**

Linett Simonsen, Tina Steinstø, Guri Verne, Tone Bratteteig. “I’m Disabled and Married to a Foreign Single Mother”. Public Service Chatbot’s Advice on Citizens’ Complex Lives. 12th International Conference on Electronic Participation (ePart), Aug 2020, Linköping, Sweden. pp.133-146, 10.1007/978-3-030-58141-1\_11 . hal-03201962

**HAL Id: hal-03201962**

**<https://inria.hal.science/hal-03201962>**

Submitted on 19 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# “I’m disabled and married to a foreign single mother”. Public service chatbot’s advice on citizens’ complex lives

Linett Simonsen<sup>1</sup>, Tina Steinstø<sup>1</sup>, Guri Verne<sup>1</sup> and Tone Bratteteig<sup>1</sup>

<sup>1</sup> University of Oslo, Gaustadalléen 23B, 0373 Oslo, Norway  
{linetts, tinaste, guribv, tone}@ifi.uio.no

**Abstract.** This paper describes a study of citizens’ chats with a chatbot of a public agency. We have analyzed chat logs and identified citizens’ lack of domain knowledge as a source of inadequate or failed chatbot responses. We identify three types of lack of domain knowledge: lack of the right vocabulary, uncertainty if a regulation fits the citizen’s situation or the “shape sorting box” problem, or citizen’s misunderstanding the regulations. The most serious failure is when a misunderstanding is not detected and corrected during the chat. The chatbot we studied is not able to make sense of badly formed questions from citizens. As implications for design we suggest making the chatbot limitations visible by not presenting it as a human-like avatar with a name. We also suggest to enable domain knowledge learning through its conversations.

**Keywords:** chatbot, domain knowledge, digitalization.

## 1 Introduction

Governmental digitalization includes the digitalization of public administration as well as the digitalization of public services, which is where the citizens meet this change. Digitalized public services are often self-service solutions where the citizens have to carry out parts of the service, hence they need to know how to use the self-service solution. Use of electronic services could decrease the need for “administrative literacy” but will increase the need for computer skills [1]. The government in Norway is therefore concerned with the “digital competence” of the population since low or no digital competence in (parts of) the population may make the digital public services inaccessible to these citizens [2]. A digital divide rooted in inaccessible digital services may have severe socio-economic consequences for the affected individuals and for society [3]. Our concern in this paper is grounded in recent studies claiming that a major reason for unsuccessful or failed use of public self-services is the lack of domain knowledge [4-7]. These studies emphasize that an important service from public call centers and staffed offices is the ability of public administration advisors to “translate” questions from citizens into the right categories for the public “machinery” to process

their inquiry. Being able to use the self-services requires some understanding of the public service system [5]. In this paper we explore how a digitalized public service handles this problem by reporting from an empirical study of a particular public service where a chatbot has been designed to be the first public encounter for a citizen who seeks information and help [8]. The study gives an in-depth description of how the chatbot functions in practice through an analysis of a set of real chatbot conversations. The paper identifies problems occurring in practice that can act as a basis for designing improved digitalized public services.

The paper first presents the public service arrangements in Norway as well as the particular service where the chatbot has been introduced. We add a brief section about chatbots in general and this one in particular before we go on to tell about our study. We have analyzed logs from chatbot chats with citizens, and the next section describes three chats each illustrating a different type of problem concerned with lack of domain knowledge. The last section discusses what sort of domain knowledge the inquiries from the public may require; our analysis of the logs is confirmed by interview data. We also discuss how chatbots may handle the problems we have identified and hint to some implications for design. The last section concludes the paper.

## **2 Background: Digital public services in the welfare state**

The public service we have studied is a part of a welfare state, which offers a range of benefits and services for citizens in various life situations, e.g., retirement pension, unemployment allowance, healthcare services and sick leave benefits, a right to education, and various forms of family benefits. The benefits are allowances to the citizen based on legally grounded arrangements. In this paper, we concentrate on family benefits, because this is the services that the chatbot we have studied is made for. If a child is born in Norway, the mother of the child will automatically receive a birth allowance. Approx. two months after the child is born, she will start receiving a monthly child benefit (until the child is 18 years old). Single mothers and fathers, who live alone with children, may also be entitled to extended child benefit and infant supplement.

### **2.1 The public service**

It is the Norway's Labour and Welfare Administration (LWA) that manages the services and benefits on behalf of the welfare state, altogether approx. 60 services and benefits. LWA meets the citizens through their 456 local offices, their Contact Center (CC) or through their digital services. The CC covers all services for the whole country and is responsible for contact with citizens through telephone, chat, and social media presence. Every year the center receives about 4.500.000 telephone calls, 700.000 written documents via net-based services (logged-in), and approx. 140.000 inquiries via chat and Facebook. The CC is not involved in the LWA's case handling: their employees only answer questions about rules and regulations, case handling response times, and payment dates. In addition, they can inform the citizen about status in her/his own case and teach citizens to use the web pages and self-service solutions.

As a part of its digitalization the LWA has expanded its digital communication and reduced the opening hours of the local offices. A chatbot has been introduced in the handling of inquiries about family benefits. A citizen who initiates a chat with LWA will first meet the chatbot before the chat can be transferred to a chat with a human. The chatbot is based on Machine Learning technology (ML) and LWA hopes that this technology combined with the large amounts of data captured by the public sector may enable automated services offering just-in-time welfare services to citizens without their active participation. The current chatbot is, however, not personal in that it does not require the citizen to log in or give any personal information.

Chatbots are expected to reduce the load on manned services, in particular calls to the call centers, by taking over the task of responding to questions from the citizens. Studying use of ML in the public sector is important for understanding how such technology functions in the public encounter [8].

## 2.2 Digitalization of the public encounter

The case handler or advisor, the street level bureaucrat who communicates directly with a citizen, represents the welfare state towards the citizens [9]. The digitalization of civil servants' work has implied that their contact with citizens and clients have changed from the "street level" contact [9], where they meet in person, to the "screen level", where data is entered into a screen, to the "system level", where decisions are made by a computer [10]. Several studies on channel choices and how they affect the relationship between the citizens and the government have been reported. Both Skaarup's study [11] of how different channels affect the citizens' feeling of control, and Ebbers et al.'s study [12] of how different channels are suitable for different purposes, show that the types of channel available can make a significant difference to the citizens. However, clients are increasingly being invited or forced to communicate with their advisors through online channels, but they also continue to use the traditional channels after adopting new digital service channels [13]. As digital technologies are used for larger parts of the communication between the client and the agency, new challenges arise. New types of skills are required from the citizen when manual services are replaced by electronic services.

There are a number of studies about what frontline service workers do and know, and many of them report that frontline service workers like the "street-level bureaucrats" [9] and the call center operators are good at "translating" callers' inquiries to internally specific categories and formats [14-16]. Several studies report that call center operators carry out knowledgeable and even emotional work in their response to a caller [17-23]. Flexible communication and emotional work are important for the operators to create a good interaction with the caller [17] and the ability to read a caller's emotional state can also be used for interpreting the emergency of the call [21].

There is a large variety of problems and issues that lead a citizen to call a public agency. Earlier studies have shown that many callers know the rules and regulations but ask if the rules apply to their particular situation; what Verne [6] calls "the shape-sorting box", i.e., fitting a real-life problem into one of the categories formed by the rules and regulations of the public agency. The callers need help to interpret the rules

and understand how they fit their own life situation [6]. However, many callers show very little domain knowledge [5, 6]. The call advisors often need to disentangle the callers' problems into smaller steps that the callers can take themselves [24]. A study of call center operators in a public agency shows that the operators responded differently to different callers depending on how they interpreted the caller's digital competence and domain knowledge: A caller appearing helpless received more guidance and practical help than callers who expressed themselves in more competent ways [25].

### 3 Chatbots

A chatbot is an automatic system to which a user can text or "chat" in real time over the Internet [26]. Today, chatbots are able to communicate with people in natural language [27]. From the very first chatbot Eliza [28, 29], most chatbots have a limited vocabulary and repertoire based on preset keywords. The Eliza program first analyzed the syntax of the user's statement then formed a response from simple pattern recognition in addition to replacing keywords. Even if chatbots are able to "seemingly reply in intelligent ways to many queries, they still cannot understand a basic sentence in natural language" [30, 31]. Moreover, Hill et al [31] found that users adapt their language by using simpler messages when interacting with a chatbot: the users they studied were very aware that they were communicating with a robot.

Today most chatbots are based on Machine Learning (ML), which is a special area within AI where statistical methods are used for identifying patterns in large data sets, to recognise patterns and produce results [32, 33]. ML is different from traditional programming where the machine performs predetermined operations. By being exposed to a number of examples or tasks with given answers, i.e., correct data, the machine is "trained". "Learning" is central in ML: the machine has to "be trained" by a set of training data and tested with a different set of test data [30, 32, 34]. The quality of the ML depends of the quality of the training data and the test data as well as how relevant these data sets are for the real data that the ML will meet when in operation. The algorithms will in principle give better results as the ML learns by processing more data. However, the development of a ML-based system like a chatbot needs to be monitored by people. A well-known example of a chatbot failure is the Microsoft chatbot Tay, which was designed to "learn" through dialogues with Twitter users and did not have security mechanisms to avoid "false teaching". Within 24 hours Tay got out of control and had to be removed from the open net due to "trolling" from the Twitter users [35].

#### 3.1 The LWA chatbot

The LWA chatbot is named Anna (a pseudonym). Anna went public in the fall of 2018 and answered chats from approximately 400 users each day in the beginning of 2019. 40 % of the users got an answer in this first encounter, 40 % were transferred to chat with a human advisor, and 20 % went to another channel (e.g., telephone) to communicate with LWA. Anna is not expected to respond to very complicated questions, as neither the technology nor the users are considered "good enough" yet.

Anna is based on ML technology and the aim is that it will learn and improve its responses through interactions with users. The ML analyses the users' questions to recognize what they are asking for. Once identified, the ML selects the corresponding response. ML is only used to identify questions: the answer the users receive to a question has been designed to adhere to the laws and regulations for that question, theme or keyword. The concrete formulation of the answers is designed by the advisors, who know what kind of replies the callers need to receive on particular questions.

In response to a question from a user, Anna pre-processes the text and classifies it as belonging to a category of questions, where all questions will match a predefined general answer. The ML performs several steps starting with correcting spelling errors and wrong words used. Then the ML identifies the question by trying to identify what the user wants to know by asking this question: the "intention" of this input. Based on the input the ML calculates which intention seems most probable. The system will match this intention with the set of predefined intentions. If the system finds the user's intention to be similar enough to a predefined intention, the predefined answer to this intention (i.e., question) will be given as the chat response. If the system does not find a sufficiently similar predefined intention, a standard fall-back message will be given, such as "Please rephrase your question in simpler terms".

Anna is constantly trained. LWA's chatbot trainers are constantly monitoring Anna, using reports generated by Anna in addition to reviewing chat conversations and feedback on Anna's performance from advisors. The chatbot trainers create training data based on this and run training sessions to improve Anna's responses. Training the chatbot means that more questions are classified, and new answers are added to the vocabulary of the system.

## 4 Method: studying chatbot logs

The main material for this interpretive qualitative case study [36] is a segment of chat logs retrieved from chats between users and the LWA's chatbot Anna from more than 6 months in 2019, each week including approx. 3000 chats. For this study we extracted approximately 8000 unique chat logs from four of these weeks: two weeks in May and two weeks in September (more detailed studies in [37, 38]). We studied documents: LWA's plans and reports as well as the chatbot developers' documentation, and we interviewed and observed advisors and chatbot trainers at two CC's units. We carried out four semi-structured interviews with three different advisors in addition to observing them as they chatted with citizens. This gave us insight into typical questions and issues that the citizens ask about. We had three semi-structured interviews with two chatbot trainers working with Anna improvements, and observed them while they worked.

We took a hermeneutic approach [36] to analyzing the data, focusing on making sense of the log data. We used Suchman's analytical framework [39] as a starting point for interpreting and understanding chat logs where difficulties had occurred. The framework (see Figure 1) distinguishes between the user's "actions available to the machine"

(i.e. input) and actions not available to the machine, and the machine’s response (“effects available to the user”) as well as the machine’s rationale or inner workings. In this paper we are concerned with the machine’s responses to the users’ actions and how Suchman’s framework allows us to include the user’s life circumstances, both those available and not available to the machine.

THE USER		THE MACHINE	
I	II	III	IV
Actions not available to the machine	Actions available to the machine	Effects available to the user	Rationale

**Fig. 1.** Suchman’s analytical framework for studying and analyzing human-machine interaction [39].

We started by analyzing each chat to see if it was successful or not. We followed up the unsuccessful ones with a more detailed analysis of the questions and responses where we coded the challenges they illustrated into different categories. In this analysis we looked for communication breakdowns and used this as a basis for understanding how the chatbot worked. Examples that we considered difficult were discussed with chatbot trainers and advisors in the LWA. In the examples given below, the questions and answers are translated by us, taking care to preserve the nuances of the legal terminology and the user’s original (mis)spelling.

## 5 Chatting with Anna

In the material we have analyzed we have found many chat logs that appear to have been successful: The user does not follow up with more questions, and our interpretation of Anna’s answer indicates that it is correct and relevant. LWA’s criteria for a successful answer from Anna is a “sufficient” answer, where the chatbot gives relevant and correct information as a response to the user input.

Some of the chat logs indicate an unsuccessful chat session. In some obvious cases, the user leaves the chat after swearing or expressing dissatisfaction with Anna’s responses, e.g., by writing “This is hopeless” before leaving. In other chats, the user does not express dissatisfaction but leaves the chat abruptly. In this section we present and discuss three examples of chat conversations that illustrate different challenges to the use of chatbots in a legal setting. We focus on lack of domain knowledge as a reason for lack of success, i.e., chat conversations where a user applies the wrong word or inconsistent terms or shows lack of knowledge about the rules and regulations. In the following three subsections, we present different chat sessions that illustrate how weak domain knowledge of the user leads to challenges in getting good answers.

## 5.1 Lack of key concept

In this category, the user formulates a question where the wrong terms are used. In this example the user asks a question about the amount of child benefit and writes the word “*child bennefit*” (translated from Norwegian with the misspelling preserved). Despite the spelling mistake, Anna recognizes that the question is about child benefits. Anna does, however, not recognize that the user asks about the amount.

**Anna:** My name is Anna and I am a chatbot.

**Anna:** I am best adapted to answering short questions. Remember to include what type of benefit you are asking about.

**Anna:** How can I help you?

**Anna:** For English chat, click here (*button*)

**User:** What is the child bennefit\* this year?

**Anna:** If you have a child under the age of 18 living with you in Norway, you may be entitled to child benefit.

**Anna:** You can receive child benefit from the month after the child is born.

**Anna:** If you are entitled to child benefit at a later point, you may receive child benefit from the month after the terms have been met.

**Anna:** Child benefit will be paid up until the month before the child turns 18. If the right to child benefit ceases earlier than this, you will receive child benefit up until the month the terms ceases to apply.

• Who is entitled to child benefit? (*button*)

• Extended child benefit (*button*)

• Payment (*button*)

• Changes in family situation (*button*)

• Apply for child benefit (*button*)

**User:** What is the rate of the child benefit?

**Anna:** Ordinary child benefit is 1054 NOK monthly per child.

**Anna:** Extended child benefit is 1054 NOK and will be in addition to the child benefit you will receive for children living with you.

**Anna:** Infant supplement is 660 NOK

• Extended child benefit (*button*)

• Infant supplement (*button*)

**User:** Thank you, Anna ☺

**Anna:** You are welcome!

Conversation continues ...

\*misspelled word

(Conversation ID: 617631 (*masked code*))

This user seems to understand that the first term s/he uses does not give the wanted answer and tries to rephrase the question using the term “rate”. The user gets a different answer, which s/he seems to be satisfied with.

## 5.2 The “shape-sorting” box

In this category, Anna cannot link the user’s question to the relevant services and benefits this user asks for. The result is that Anna does not recognize what the user asks for, and therefore cannot help.

**Anna:** My name is Anna and I am a chatbot

**Anna:** I am best adapted to answering short questions. Remember to include what type of benefit you are asking about.

**Anna:** How can I help you?

**Anna:** For English chat, click here (*button*)

**User:** Going to have a child, what do I need to apply for?

**Anna:** I’m sorry, but I don’t understand what you are asking.

**Anna:** I will understand it better if you write short questions and which benefit it is regarding.

**Anna:** Can you ask the question again?

• Need more help (*button*)

• No (*button*)

**User:** How do I receive money when I am expecting a child?

**Anna:** Which day the payment arrives, depends on which benefit (it regards).

**Anna:** You can see your payments from LWA at “Your payments” at <LWA.website>, but they do not show up before they have arrived in your account.

<lists info about payments>

(Conversation ID: 679785)



The question formulated by the user in this example indicates weak domain knowledge, as the user does not know what the different services and benefits are called, or who is entitled to what. We call this a problem with “the shape-sorting box”: matching a citizen’s life situation to the various services and benefits.

The user leaves the chat without asking further questions. If this user had asked explicitly to be transferred to chat with an advisor, it would have been visible in the chat log. S/he may have changed the contact to another channel or given up without the chat log giving any indication.

### 5.3 A logical/legal flaw

In this category, the terms used indicate a weak understanding of the rules and regulations for LWA’s benefits and services. In this chat, the user asks the question “can a young disabled person married to a foreign single mother receive child benefit and child supplement”. The question contains the legally inconsistent phrase “married to a ... single mother”: the rules are that if you are married you are not categorized as a single parent. Anna’s response concerns benefit for single parents, which does not match the user’s stated life situation. The user follows up with asking whether a “foreign single mother” can receive child benefit and child supplement. Anna does not recognize that the user expresses a legal inconsistency and responds by providing more information about the rights of single parents.

**Anna:** My name is Anna and I am a chatbot.

**Anna:** I am best adapted to answering short questions. Remember to include what type of benefit you are asking about.

**Anna:** How can I help you?

**Anna:** For English chat, click here (*button*)

**User:** can a young disabled person married to a foreign single mother receive child benefit and child supplement

**Anna:** If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children.

**Anna:** The extended child benefit is child benefit for one more child than what you actually are living together with.

**Anna:** You have to apply for an extended child benefit, it will not be automatically granted.

- Who is entitled to child benefit? (*button*)

<lists info about extended benefit>

**User:** Entitled to extended child benefit? (*action*)

**Anna:** If you are a single mother or father you can receive extended benefit if you are living alone with the child and:

- you are separated, divorced or widowed

<continues to list criteria for extended benefit>

**Anna:** If you do not have common children with the person you are living with, we do not need documentation that you are not living together. It could be:

- Declaration of breakup

- Separation permit

- When does the right to extended child benefit expire? (*button*)

- Apply for extended child benefit (*button*)

- Shared child benefit (*button*)

**User:** When does the right to extended child benefit cease? (*action*)

**Anna:** Extended child benefit will cease if:

- You marry

- You have children with a new cohabitant

- You have been cohabitant with a person you do not have joint children with for twelve months

- You move together with the father of the child

**User:** can a foreign single mother receive child benefit and child supplement

**Anna:** If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children.

**Anna:** <lists criteria for extended benefit>

**User:** Payment (*action*)

**Anna:** Child benefit will be paid out the last day of the week each month. Child benefit is not liable to tax.

**Anna:** Remember that you can see your payments from LWA by logging into ‘Your LWA’.

**Anna:** At <LWA.website> you can find an overview of payment dates for child benefit.

<lists payment info>

**User:** Can foreign single mother receive child supplement

**Anna:** In some cases, the mother may receive the entire period of child benefit

**Anna:** She can be alone with the care or live together with the child's other parent that have not yet served up the right to child benefit.

- I am caring for the child alone (*button*)
- I live together with the father/co-mother (*button*)
- I mean child benefit (*button*)
- I mean transitional benefit (*button*)

**User:** I mean child benefit (*action*)

**Anna:** If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children.

**Anna:** The extended child benefit is child benefit for one more child than what you actually are living together with.

**Anna:** You have to apply for an extended child benefit, it will not be automatically granted.

<lists info about extended benefit>

(Conversation ID: 670860)

Nowhere in this conversation does Anna explain that you cannot be categorized as a single mother if you are married. As the user continues to ask about benefits for single mothers, this misunderstanding is not detected, and Anna continues to answer about benefits for single mothers without explaining that the mother in question may not be eligible for such benefits. The user selects the button “I mean child benefit” and is given a response that Anna has already provided earlier in the conversation. Since Anna's reply contains the buttons “I mean child benefit” and “I mean transitional benefit”, it may look as if the system recognizes the user's weak domain knowledge. However, the answers the user receives may be completely irrelevant if the mother in question actually is married. In the worst case, this user is misled by the chatbot into believing that the mother he is married to is eligible for benefits for single mothers, which she is not.

## 6 Discussion

### 6.1 Chatbots as a frontline in public services

Previous research about call centers has shown that the operators do more than merely answer questions about facts and information. The operators' expertise and knowledge about their work add value to their clients' inquiries [19]. Whalen et al. [23] describe operators' work by the seemingly contradictory notions of improvisation and choreography, indicating that their work routines are craft-like practices where the operator adjusts his/her response to the situation-at-hand: the actual, concrete inquiry from the client. Verne's study of call center operators in a public agency showed that operators helped callers to formulate their inquiries in the relevant and correct terms as well as providing problem-solving steps that the caller could perform to solve the problem [25].

There is no creative translation capacity in a chatbot: the chatbot only compares the words given to it through the user's input and matches these to the predefined keywords – the only pre-processing done by the chatbot is spell check. The cases above show that the chatbot often does not forgive spelling errors or use of wrong words by clients. This makes the chatbot a very strict first encounter with the welfare agency for many citizens. The chatbot becomes a representative for LWA, hence the welfare state may appear rigid when it does not provide information that the citizen is not aware of that s/he should ask for.

## 6.2 Domain Knowledge

Our study shows different ways that the lack of domain knowledge affects the citizens conversations with the welfare state through chatbots. We have identified three types of domain knowledge that may result in unsuccessful communication with chatbots like Anna if missing.

The first type of domain knowledge is the language: the correct terms. The user in our first case misspells a (correct) term and also uses an everyday word (“amount”) that Anna does not recognize. Lack of knowledge about the correct terms is a well-documented problem that goes deeper as knowledge about the correct terms often depends on knowledge about the phenomenon that the terms refer to, i.e., the rules and regulations of the public service. Our interviews with advisors confirmed that many citizens contact them using everyday language. “It is easier for them to use the chat and say ‘Hi, I am new in this. And I wonder about so-and-so’. Maybe they don’t know where to start. These are the people I think contact us on the chat,” said one of the advisors. Another advisor confirmed this by stating that “My experience is that many people ask like ‘Hi, I plan to have kids, my due date is soon, my due date will be like so-and-so. What should I take care of? How do I apply?’” The tax call advisors in Verne’s study [6] recognize weak domain knowledge by wrong or inconsistent terms used by the callers. A study among immigrants in Norway showed that even if they had lived in Norway for several decades and spoke Norwegian well, the specific terms of LWA were difficult to understand and master because the terms had no equivalents in their mother tongue [40]. This also holds for native Norwegians without competence in LWA’s services: the advisors told that they “normally use an oral language style to make it easier to understand for the users” in the chats and that they “break the information down to make it more understandable”. Another advisor told us that “it is very rare that you can use the standard answers”, so the answers have to be adjusted. “I assess every chat and adjust the information not too much and not too little to the concrete chat.” Moreover, the advisors emphasized that the users ask in many different ways: “There are incredibly many nuances and incredibly many ways of solving things”.

The second type of domain knowledge also seems to be a language problem, but we categorize it as a “shape sorting box” problem. The user asks if s/he is eligible for a particular benefit. This problem is discussed by Verne [6]. Her study of the tax authorities’ call center concluded that many of the callers asked how their particular life situation fitted with the rules – even callers with good knowledge about the rules wanted to have their interpretation of the laws and regulations confirmed from the official view of the tax authorities. This was supported by the advisors: “Very often they want a confirmation because it is complicated regulations and many nuances” and “the users contact us to make sure they have done things right”. The advisors also said that “much of the information [the users ask for] you can find at the <LWA.website>, but we understand that they chat with us. Because it is a lot of information. And they have read it, but they have not really understood it properly”. They also say that “The application [for benefits] has become a lot easier, but people don’t stop contacting the LWA”. One advisor explained that “A challenge in LWA is that things overlap a lot. There are no strict borders, parental things flow into [other topics] ... There are a lot of terminologies

that are similar, but not quite the same, but almost the same: child benefit, child allowance ...”. Anna is not able to answer questions about the “shape sorting box” because the chatbot only answers general questions, referring to general rules and regulations. Anna helps people find general information but cannot relate to the actual situation of a person.

The third type of domain knowledge is a misunderstanding of the rules that looks like a “shape sorting box” question. However, as the chatbot does not detect this misunderstanding, there is a danger of mis-informing the caller, who may get an unpleasant surprise if s/he has calculated with a benefit it later turns out that s/he is not entitled to. The advisors told us that “Many people do not understand what is written at <LWA.website>”. Combined with what the advisors told about adjustment of their language to the individual user and the many nuances in both questions and answers, we find that the human chat offers what call centers normally do: “I answer on their level”.

### 6.3 Implications for design

In this study, we have found several challenges in citizens’ chat with Anna that suggest shifting the focus in chatbot design from a focus on the chatbot end of the interaction to taking the whole chat conversation into consideration (in line with [41, 26, 43]). Our analysis of the chatbot’s responses to the users’ questions (cf. Figure 1, [39]) has of course made use of our (human) interpretation of the content of the questions and what they mean as well as the assumed value and appropriateness of the chatbot’s answers. Based on this analysis we suggest some implications for the design of the chatbot end of those conversations. As users’ lack of domain knowledge may cause miscommunication and communication breakdowns, we encourage chatbot designers to explore the possibility of helping users gain domain knowledge, for instance by providing cues or basic domain information as part of the chatbot communication.

Our analysis indicates that users’ expectations were out of step with the practical realities of chatbot capabilities, causing frustration and dissatisfaction on the user side. One reason for the high expectations may be that LWA’s chatbot Anna is presented with a human name and a human-like avatar. The chatbot logs we studied contained anthropomorphic trigger responses such as “Aww, now you are making me blush” and “I understand very well what you are saying”. Our analysis suggests that not all users understand that they communicate with a computer, and we suggest designing the chatbot in a way that does not try to impersonate a human, but clearly reveal its non-human qualities. Programmed trigger responses may set unrealistic expectations framing the ongoing user experience [17, 27, 42], hence we suggest a careful approach when implementing such features.

Reflecting on the introduction of Anna, one of the advisors say that “It has been a pronounced change – there are fewer conversations on [the human] chat but I do not know if more people get help”. As of today, LWA’s chatbot can only be used as a non-logged in service. However, it is possible to imagine a future where users can communicate with the chatbot as logged in users. In this case the system may have access to personal information about the users, e.g., if they are married or have children. This

may affect how the chatbot system will be able to help users in their particular situations. The chatbot can, e.g., check if the marital status mentioned in the question is in line with what is registered in their databases and respond appropriately.

The fact that the chat is the first encounter with the welfare state for the citizen makes it important that the chatbot's answers are to be trusted. We saw a small sign that the trust is fragile in a chat conversation with a human, where the user asks a question and the human chat advisor recommends a form and provides the link to the form. The user responds by asking "Is the form you linked to the right one?" The advisor had to explain the reasons for recommending this particular form. Chatbot encounters that make users lose trust in the welfare state may be a serious cost of the benefits of digitalization.

## 7 Concluding remarks

In this paper we have explored how a digitalized public service in the form of a chatbot handles citizens' inquiries about family benefits. By analyzing logs from chatbot chats with citizens we have shown that chats where citizens lack domain knowledge may result in inadequate responses by the chatbot. Lack of domain knowledge is exemplified as lack of the right vocabulary, uncertainty if a regulation fits the citizen's situation – the "shape sorting box" problem, or misunderstanding the regulations. We find that the lack of adequate response from the chatbot is particularly problematic if a misunderstanding that the citizen has is not detected and addressed during the chat. Our studies of real chats with a chatbot show that the chatbot is not able to interpret and "translate" badly formed questions from citizens into the right categories for further advancement of their inquiry. We suggest making the limitations of the chatbot's capabilities more visible by presenting it as a computer rather than a human-like avatar as well as expanding its ability to explain and teach citizens relevant domain knowledge through its conversations. Lack of domain knowledge makes the user unaware of inconsistencies and limitations in the chatbot's answers.

## Acknowledgements

We thank LWA Contact Center and its leader for giving us access to the chat logs and for allowing us to interview the advisors and chatbot trainers.

## References

1. Grönlund, Å., Hatakka, M., Ask, A.: Inclusion in the E-Service Society – Investigating Administrative Literacy Requirements for Using E-Services. In: Electronic Government: 6th International Conference (EGOV 2007), 4656/2007. Springer Berlin, pp. 216–227, (2007).
2. Carretero, S., Vuorikari, R., Punie, Y.: The Digital Competence Framework for Citizens. Publications Office of the European Union (2017).
3. Van Deursen, A. J., Van Dijk, J.A.: The First-level Digital Divide Shifts from Inequalities in Physical Access to Inequalities in Material Access. In: *New Media & Society* 21 (2), pp. 354–75 (2019).

4. Breit, E., Salomon, R.: Making the Technological Transition – Citizens' Encounters with Digital Pension Services. In: *Social Policy & Administration*, 49, pp. 299–315 (2014).
5. Madsen, C. Ø., Kræmmergaard, P.: The efficiency of freedom: Single parents' domestication of mandatory e-government channels. In: *Government Info. Quarterly*, 32 (4), pp. 380–388 (2015).
6. Verne, G.: The winners are those who have used the old paper form. On citizens and automated public services. Doctoral dissertation, Dept. of Informatics, Univ. of Oslo, (2015).
7. Verne, G., Bratteteig, T.: Do-it-yourself services and work-like chores: on civic duties and digital public services. In: *Personal Ubiquitous Comput.*, 20 (4), pp. 517–532 (2016)
8. Lindgren, I., Madsen C.Ø., Hofmann, S., Melin, U.: Close Encounters of the Digital Kind: A Research Agenda for the Digitalization of Public Services. In: *Government Information Quarterly* 36(3), pp. 427–36 (2019).
9. Lipsky, M.: *Street-Level Bureaucracy. Dilemmas of the Individual in Public Service*. Russel Sage (2010).
10. Bovens, M., Zouridis, S.: From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. In: *Public Administration Review*, 62, pp. 174–184 (2002).
11. Skaarup, S.: The Mediation of Authority. In: *scss.tcd.ie*, pp. 1–32 (2012).
12. Ebbers, W.E., Pieterse, W.J., Noordman, H.N.: Electronic government: Rethinking channel management strategies. In: *Government Info. Quarterly*, 25 (2), pp. 181–201, (2008).
13. Madsen C. Ø., Hofmann S., Pieterse W.: Channel Choice Complications. In: *International Conference on Electronic Government 2019* Springer, pp. 139–151 (2019).
14. Wynn, E.: Office conversation as an information medium. Doctoral dissertation, University of California, Berkeley, Department of Anthropology (1979).
15. Suchman, L., Wynn, E.: Procedures and problems in the office. In: *Office: Technology and People* 2, pp. 133–154 (1984).
16. Bowker, G. C., Star, S.L.: *Sorting Things out: Classification and Its Consequences*. Cambridge, Massachusetts: MIT (1999).
17. Maass, S., Rommes, E.: Uncovering the Invisible: Gender-Sensitive Analysis of Call Center Work and Software, *Gender Designs IT: Construction and Deconstruction of Information Society Technologies*. In: Zorn, I. et al. (eds.) *Gender Designs IT*. Wiesbaden: VS (2007).
18. Martin, D., O'Neill, J., Randall, D., Rouncefield, M.: How Can I Help You? Call Centres, Classification Work and Coordination. In: *Journal of CSCW* 16 (3), pp. 231–264. (2007)
19. Muller, M. J.: Invisible Work of Telephone Operators: An Ethnocritical Analysis. In: *Journal of CSCW* 8 (1-2), pp. 31–61 (1999).
20. Nyberg, D.: Computers, Customer Service Operatives and Cyborgs: Intra-actions in Call Centres. In: *Organization Studies* 30 (11), pp. 1181–1199 (2009).
21. Svensson, M.: Routes, Routines and Emotions in Decision Making of Emergency Call Takers. Doctoral Dissertation, School of Management, Blekinge Institute of Technology (2012).
22. Tjora, A.H.: The technological mediation of the nursing-medical boundary. In: *Sociology of Health & Illness* 22 (6), pp. 721–741 (2000).
23. Whalen, J., Whalen, M., Henderson, K.: Improvisational choreography in teleservice work. In: *The British Journal of Sociology* 53 (2), pp. 239–258 (2002).
24. Bratteteig, T., Verne, G.: Conditions for Autonomy in the Information Society; Disentangling as a Public Service. In: *Scand. Journal of Information Systems* 24, pp. 51–71 (2012).
25. Verne, G.: Two faces of autonomy: Learning from non-users of an e-service. In: *System, signs and actions*, 8 (1), pp. 6–24 (2014).
26. Brandtzaeg, P., Følstad, A.: Chatbots: Changing User Needs and Motivations. In: *Interactions*, 25 (5), pp. 38–43, ACM (2018).

27. Liao, Q.V., Hussain, M.M., Chandar, P., Davis, M., Khazaen, Y., Crasso, M.P., Wang, D., Muller, M., Shami, N. S., Geyer, W.: All Work and No Play? Conversations with a Question-and-Answer Chatbot in the Wild. In: CHI 2018, pp: 3.1-3.13, Montreal, QC, Canada (2018).
28. Weizenbaum, J.: *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: Freeman (1976).
29. Shevat, A. *Designing bots: creating conversational experience*. Beijing (2017).
30. Holmquist, L. E.: Intelligence on Tap: Artificial Intelligence as a New Design Material. In: *interactions*, 24 (4), pp. 28–33, ACM (2017).
31. Hill, J., Ford, W.R., Farreras, I. G.: Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. In: *Computers in Human Behavior*, 49, pp. 245–250 (2015).
32. Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. In: *Science*, 349 (6245), pp. 255–260 (2015).
33. Russell, S., Norvig, P.: *Artificial Intelligence: A modern approach*, Pearson, Boston, (2010).
34. Broussard, M.: *Artificial unintelligence: how computers misunderstand the world*. Cambridge, MA, MIT Press (2018).
35. Hendler, J., Mulvehill, A.: *Social Machines: The Coming Collision of Artificial Intelligence, Social Networking, and Humanity* (1st). Berkeley, CA, USA: Apress (2016).
36. Myers, M.D.: *Qualitative Research in Business & Management*, 2<sup>nd</sup> edition, Sage Publications (2013).
37. Simonsen, L.: Når brukerdialogen automatiseres – hva blir vanskelig? En kvalitativ studie av sekvensen av handlinger mellom menneske og chatbot. Master thesis, Dept. Of Informatics, University of Oslo (2019).
38. Steinstrøm, T.: Bruk av chatbot i praksis. En kvalitativ studie av utfordringer ved bruk av chatbot i offentlige tjenester. Master thesis, Dept. Of Informatics, University of Oslo (2020).
39. Suchman, L. *Human-machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge: Cambridge UP (2007).
40. Nguyen, T.H.: «Jeg blir så sliten i tankene». Innvandreres bruk av offentlige nettbaserte tjenester. Master thesis, Dept. of informatics, University of Oslo (2019).
41. Hageback, N.: *The Virtual Mind: Designing the Logic to Approximate Human Thinking*. Boca Raton, FL: CRC Press (2017).
42. Luger, E., Sellen, A.: Like Having a Really Bad PA: The Gulf Between User Expectation and Experience of Conversational Agents. In: CHI 2016, pp: 5286–5297 (2016).
43. Følstad, A., Brandtzæg, P. B.: Chatbots and the New World of HCI. In: *interactions*, 24 (4), pp. 38–42 (2017).