

12-31-2022

How Can I Help You? A chatbot's answers to citizens' information needs

Guri B. Verne
University of Oslo, guribv@ifi.uio.no

Tina Steinstø
University of Oslo, tina-ste@hotmail.com

Linett Simonsen
University of Oslo, linettsimonsen@gmail.com

Tone Bratteteig
University of Oslo, tone@ifi.uio.no

Follow this and additional works at: <https://aisel.aisnet.org/sjis>

Recommended Citation

Verne, Guri B.; Steinstø, Tina; Simonsen, Linett; and Bratteteig, Tone (2022) "How Can I Help You? A chatbot's answers to citizens' information needs," *Scandinavian Journal of Information Systems*: Vol. 34: Iss. 2, Article 7.

Available at: <https://aisel.aisnet.org/sjis/vol34/iss2/7>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Scandinavian Journal of Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

How Can I Help You?

A chatbot's answers to citizens' information needs

Guri B. Verne, Tina Steinstø, Linett Simonsen and Tone Bratteteig

University of Oslo

guribv@ifi.uio.no, tina-ste@hotmail.com, linettsimonsen@gmail.com,

tone@ifi.uio.no

Abstract. AI-based chatbots are becoming an increasingly common part of the front-line of public services. Through natural language, users can write simple queries to a chatbot which answers with appropriate information. We have investigated how a public chatbot operates in actual practice and how it answers the citizens' questions about the rules and regulations for welfare benefits. We use the concept of citizens' information needs to determine the quality of the chatbot's answers. Information needs are often not formulated from the start as answerable questions. We analyse logs from chat sessions between the chatbot and the citizens, and focus on problems that arise, e.g., that the chatbot gives irrelevant answers or omits important information. The paper shows how the inner workings of the chatbot shapes the answerable questions. We conclude that responsible use of AI (such as chatbots) is a matter of design of the overall service and includes acknowledging that the AI itself can never be responsible.

Key words: chatbot, citizens, information needs, human-machine interaction, transparency, responsible AI.

1 Introduction

A chatbot is very often the first thing you encounter when you contact an organization. Many organizations have delegated the first line of communication with customers to chatbots that can direct the customers to a website with the information they ask for. Digitized self-services save time and money for the organization and enables a fast 24 hours a day, 7 days a week service to the customers. Also public and governmental institutions report large savings when chatbots and web-based self-services take over much

Accepting editor: Elena Parmiggiani

of the communication with the public. The number of citizens using digitalized public services is increasing. However, many public institutions still have to offer conversations with human advisors (physical meetings, phone and chats) in cases when automated chatbots do not provide satisfactory answers. In this paper we report from a case study investigating if citizens using a public service chatbot get the information they need.

A chatbot is an example of *artificial intelligence* (AI) technology, which is an increasingly common element in digitalized private organizations and government institutions. AI is used to improve the services and automate work processes. Today, AI includes *machine learning* (ML) that makes use of data collected from use and users resulting in data-driven, automatic systems that are more complex and difficult to understand (Burrell, 2016; Jordan and Mitchell, 2015; Lyytinen et al., 2020; Muller et al., 2019; Russell and Norvig, 2010). AI transforms the services and it transforms the work carried out by the human service workers as well as the work carried out by the users of digitalized services (Verne and Bratteteig, 2016).

We report from a study of citizens' chatbot conversations with a public welfare institution. The chatbot answers the citizens' questions or directs them to the appropriate web site where the topic they ask for is explained. The chatbot in our study is employed by a national public welfare administration (WA). WA delivers social benefits to the public, i.e., benefits for unemployment, disability, parental and child welfare, sick leave etc. For a citizen, communication with the WA normally concerns investigating if the citizen qualifies for a particular benefit. The rules and regulations for being eligible for the benefits are often complex and have exceptions, and the process of applying is difficult to navigate for citizens not knowledgeable in the welfare system. As a public welfare institution, the WA is obligated to inform citizens about their rights and duties and to support those in need. It is responsible for giving correct and relevant information to citizens' inquiries. On their website, the WA presents the chatbot Anna (a pseudonym) as the first contact point for the citizens. Anna handles an increasingly larger part of the WA's communication with the citizens. Its answers to the citizens' inquiries may affect how they can handle their own life situation and move forward.

Delegating to a chatbot to present public information about welfare benefits comes with special challenges and responsibilities, especially when the chatbot's responses will inform the citizen about rules and regulations that will affect the citizens' life situation. Lack of an adequate response from the chatbot can be particularly problematic if a misunderstanding from the citizen is not detected and addressed during the chat session. Studies show that errors are common in AI-infused systems (Amershi et al., 2019). Responsible use of AI should ensure "effective and responsible outcomes" (Mikalef et al., 2022, p. 7). Citizens who trust an answer to be correct that is actually irrelevant to

their situation may believe that they are entitled to benefits when they are not, or they may not get sufficient information about the benefits they can apply for. Making sure that citizens understand how to ask questions and assess the answers from the chatbot, is therefore important (Simonsen et al., 2020).

The chatbot Anna is the first point of contact with the WA. There is a waiting line for communication with a human advisor in the chat or on the phone. For citizens using the WA, there is no other welfare agency to contact if they do not like the services that the WA provides. Our overall question in this paper is if citizens using a public chatbot get the information they need and if not, why? We have analysed real chatbot conversations with a public institution (i.e., the WA) and have looked for conversations where citizens express that they are not satisfied with the chatbot's answers or request to communicate with a human advisor. We have particularly looked for conversations where the citizens do not seem to discover that the answer is not good in the sense that it does not fully inform the citizens about their legal rights. Moreover, we are interested in how and why the technology produces its answers; that is, how well the chatbot matches citizens' questions with answers. We try to answer this question by investigating how the chatbot helps the citizen to formulate questions that can receive good answers and also include how the chatbot's ML technology interprets the citizens' questions. The discussion of how well the chatbot answers citizens' questions contributes to describing some of the challenges that public service institutions meet when aiming to use AI in a responsible way.

The paper is organized as follows. In the next section, we discuss chatbots and some challenges when putting a chatbot into actual use. Section 3 introduces a theory about information needs from information science, describing four levels of information needs. We use this framework to analyse the chat conversations. In Section 4, we describe the WA's chatbot Anna and its inner workings. Section 5 reports from the research methods we have used. We describe the analytical framework we have used for including the technology's operations into the analysis of the chats. In 6, we present and analyse a selection of six chat sessions in detail: we have selected examples that can show how the chatbot responds to citizens inquiries. In Section 7, we discuss the analysis while Section 8 concludes the paper.

2 Chatbots in use

A chatbot is a computer system that a user can text or 'chat' with over the Internet in real time (Brandtzæg and Følstad, 2017). Most people use chatbots for productivity reasons, as an easy way to get answers instead of making a phone call or reading through

a lot of text (Brandtzaeg and Følstad, 2017). However, constructing a good chatbot that meets the users' expectations and gives adequate responses is challenging. Several chatbots have failed to realize the expectations of developers and users, and for some chatbots, humans are filling in responses behind the scenes to camouflage insufficient chatbot performances (Grudin and Jacques, 2019).

The first and perhaps most well-known chatbot was Eliza, designed by Joseph Weizenbaum to demonstrate the limits of AI (Shevat, 2017; Weizenbaum, 1976). Eliza was a computer program simulating a psychologist and appearing to be able to understand and communicate with the user in natural language. Eliza was based on simple scripts for language analysis where the output was based on a reformulation of keywords in the input. Despite knowing that they communicated with a computer program, many people used and got involved emotionally with the Eliza chatbot (Weizenbaum, 1976). Today, methods for ML have enabled a development of more advanced chatbots that can learn from historical data (Jordan and Mitchell, 2015). ML involves training a computer model by a set of training data and testing it with a (different) set of test data (Broussard, 2018; Holmquist, 2017; Jordan and Mitchell, 2015). The quality and relevance of the data used for training and testing is important for the quality of the responses from the ML algorithm when in operation.

Chatbots are used for many different purposes, such as providing information, customer services, entertainment or even socializing. Følstad and Brandtzaeg (2017) studied why users choose to use a chatbot and find that they help users to obtain timely and efficient assistance or information in addition to being used for social and entertainment reasons. Chatbots for customer service need to be trusted by the users (Følstad et al., 2018). Research on chatbots from a chatbot provider's perspective includes studies of how a chatbot meet the users' expectations (Kosielnik et al., 2019; Zamora, 2017) or how it helps or stimulates students (Dibitonto et al., 2018; Fryer et al., 2017). Many studies of chatbot design are carried out from the provider's side with a focus on how to drive profit and customer satisfactions (Adam et al., 2020). For chatbots used in education there are so far hardly any studies investigating the information needs of the learners with respect to the chatbots nor if chatbots address these needs sufficiently (Wollny et al., 2021).

There are also a few studies analysing how a chatbot operates in practice from a user's perspective. Chatbot users are, for example, often aware that they communicate with a robot and adapt their language by using simpler messages (Hill et al., 2015), possibly losing some of the precision in their requests. In a study of conversations with a customer chatbot provided by a telecommunication company, most of the feedback given by the users was negative (Akhtra et al., 2019). The study used data mining techniques

and feedback mechanisms in the chatbot window to assess the users' topics of interest and if the users were satisfied with the chatbots' responses. A user's positive evaluation of the experience of using a chatbot may extend beyond satisfied information needs (Liao et al., 2018), e.g., getting a positive answer. For a private company, user satisfaction may be important for their sales.

There are challenges for using AI in the public sector concerned with social and ethical aspects as well as legal conditions, responsibility, and accountability (Wirtz et al., 2019). User satisfaction is only one criterion, as correctness and relevance in the chatbot's answers may be more important. Chatbots for public institutions are expected to transform communication between citizens and government (Androutsopoulou et al., 2019). We have not found studies about how the chatbot meet citizens' information needs about welfare benefits. This is a domain where the chatbot will need to answer with correct and relevant information and where erroneous or missing information may be detrimental for the citizens who need to apply for benefits.

3 Theoretical frameworks

Our analysis builds on two theoretical sources, both addressing aspects of the conversation between a chatbot and a human user. The first framework we use is adapted from Suchman's book about human-computer interaction (1987; 2007). Her simple framework can be used for analysing details of an interaction to find out how and why it works or not. In our second framework we draw on information and library science, where people's information behaviour has been studied for years. We refer to a classic framework by Taylor (1968) characterizing how people work with expressing their information needs.

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

Figure 1. Suchman's analytical framework (1987; 2007) for studying and analysing human-machine interaction.

3.1 Suchman's framework for Human-Machine Interaction

Suchman's analytical framework (1987; 2007) focuses on the human input to the machine and the output that we can see or experience understood as a sequence of actions between the human and the machine, where each of them responds to the other's actions, see Fig 1. Its core are the actions and their effects that are available to the user and the machine, respectively.

What the user does that the machine can detect, e.g., pushing buttons, are 'actions available to the machine'. The machine's responses that the user can experience are 'effects available to the user'. What the user does that is not recognized by the machine, e.g., talking or carrying out irrelevant actions, are 'actions not available to the machine' as the machine cannot detect and respond to such actions. Suchman's framework also includes a fourth column indicating the 'rationale' for why the machine did what it did, which often is not what the users expect. Suchman did not take the technology 'inside' the machine into account, but gave an explanation of the rationale behind the actions of the machine. As an anthropologist Suchman did neither aim to understand nor explain the technology that produced the responses from the machine. In this paper we have extended her framework to include technical explanations in addition to the rationale behind the machine's responses in column IV.

This simple framework was made to analyse when and how the interactions between the human and the machine came out of sync, i.e., when the machine did not respond adequately to the user's input. The original machine described in Suchman (1987; 2007) was a script-based photocopier where the user interacted by pushing buttons or loading paper. We have extended column II in Fig 1 to also include textual information given by the user in addition to the actions carried out.

The framework helps us match the response from the chatbot with the question from the user, as it allows us to focus on how information given by the user is reflected (or not) in Anna's reply. This has been important for understanding if the chatbot's reply addresses all the information given by the user in an adequate way when they formulate their request for information. To analyse the information needs in more detail we have used the four levels of information needs suggested by Taylor (1968).

3.2 Taylor's theory of Information needs

Information needs are often complex and difficult to formulate precisely. However, precise questions are a key to get relevant answers. In a classic and influential paper, Taylor (1968) describes how a need for information starts from a feeling that iteratively develops into a precise formulation of a question that can be answered by an information

system. Taylor's original 1968 paper concerns the search for literature in a university library, but it has attracted interest in disciplines outside of library science. Today his theory is used in, e.g., computer science, medicine, engineering and education (Chang, 2013).

Central in Taylor's paper is that a person seeking information "tries to describe for another person not something he knows, but rather something he does not know" (1968). A request for information does not come into this person's mind fully formed, instead the need for information can begin as an unrest, a feeling that one needs to find out something. Perhaps a person is experiencing a new life situation and wonders if there are welfare benefits for this new life situation. At first, the person does not consciously know the precise need for information such that a precise search request can be formulated. First, the person formulates the request in relatively open and imprecise terms, but through a dialogue with a specialist in the field the request will become more precise. To receive a relevant response from an information system, the person needs to formulate and perhaps reformulate the request in terms adequate for this particular system (Taylor, 1968).

Taylor's classic paper (1968) focuses on the people seeking information, not the systems and tools for information seeking (Tyckoson, 2015). A person seeking information receives help in conversation with a specialist to refine the request, often in several iterations, so that the request can be adapted to an information system and receive a precise answer. During this process of question negotiation, information specialists help transforming actual information needs into precise formulations (Chang, 2013).

Taylor (1968) describes four levels of information needs which represent iterative reformulations of a request for information. A request develops from an unformulated need on level 1, through a first conscious expression on level 2, to a well-formulated question that receives a relevant answer on level four. In Taylor's original paper, these levels describe the dialogue between a librarian and a library visitor who needs to find literature in the library. The request for information is negotiated with the librarian's knowledge about the library's information system (and the books) and the formulation of the question is refined through this dialogue. The dialogue is seen as a process of question negotiation, where the library visitor's open formulations of information needs are transformed into more precise expressions that the information system can handle (Chang, 2013).

In our use of Taylor's theory, the chatbot fills the roles of both the specialist and the information system. The WA chatbot Anna is able to answer citizens' requests for information about welfare rules within several domains (but not all). The citizens cannot be expected to know much about the welfare rules before they engage with the chatbot;

the welfare area is often new to them, e.g., if they become sick or parents for the first time.

In our analysis, we categorize the requests to Anna into the appropriate level of expressing an information need. We use the term ‘request’ instead of ‘question’ or ‘inquiry’ as many citizens using the chatbot in our material do not write a proper question; often they merely state a need or describe a life situation. Below we present the four levels of information needs from Taylor (1968) and how we interpret them in this study.

The first unformulated wish for information is seen as information need on level 1, written as Q1. This is the actual, but unexpressed need for information experienced as a bodily sensation. A person may want to find out more about welfare benefits, normally because of a new life event. Taylor called this the *visceral need*. Level Q1 represents an unexpressed information need; hence, we did not classify any questions for Anna to be on this level.

The first attempts to formulate an information need to Anna will often be on level Q2, which is the conscious description of the information need for oneself. The citizen is aware of an information need and makes a first formulation of it, perhaps expressed as a simple question or some keywords to the chatbot. Taylor called this the *conscious need*. In our study, an example of a request categorized as a Q2 is a citizen describing his or her life situation in a simple manner, e.g., ‘pregnant student’ or ‘I need money’. An attempt from the citizen to formulate a request to Anna using welfare terminology, will in our interpretation be on level Q3. However, it may also be a reformulation of a previous Q2 request. Taylor called this the *formalized need*. In our material, an example request is ‘When do I get money?’

A question for Anna that receives a relevant and correct answer will be on level Q4. In some of the chat sessions, the citizen refines a question that did not give a relevant response and the new formulation gives a better response. Taylor called this the *compromised need*, as this expression will be the result of question negotiation in a dialogue with the domain expert and adapted to the information system, in our case the chatbot. In our material, an example will be ‘What is the child maintenance payment for a six-year-old child?’ Taylor makes a point that these four levels are to be seen “only as convenient points along a continuum” (ibid., 182), and we have followed his advice in our interpretations of the Q-levels from a rich material of requests.

4 About the WA chatbot Anna

Governmental institutions represent the state, and it is particularly important that their services are in line with the legislation and that the decisions they make are fair. Public

institutions have an obligation to give information about rules and regulations in their field to the citizens. Traditionally, civil servants in the front line have represented the public service, constituting the public policy by carrying it out in practice in direct contact with citizens. In many public services, a large part of the work of these so-called “street-level bureaucrats” (Lipsky, 2010) is to customize general rules and legislation to the individual citizen, hence, the opportunity to exercise discretion is crucial (Verne et al., 2022). When a chatbot is the citizen’s first encounter with a public service, the chatbot comes to represent the public service towards the citizen. The chatbot represents the public service institution and should fulfil its obligations and responsibilities, which is a measure of how well the chatbot functions in practice.

4.1 Anna in practice

The WA chatbot we have studied, Anna, was among the first chatbots from the Norwegian government. Anna went public in the fall of 2018. At the time of our study it was limited to only answering questions about parental and child benefits¹. By early 2019, Anna handled approximately 400 citizen requests per day. According to the WA, when the chatbot represented the citizens’ first encounter with the agency, 40% got their question answered by Anna, 40% of the citizens were transferred to chat with a human advisor, and 20% were transferred to another communication channel (e.g., telephone). In 2020, the WA registered 97 million visits to their webpage, and received 3.2 million phone calls to their contact center, who runs the chatbot. In a 2020 user survey carried out by the WA, more than 60% answered ‘no’ to the question if Anna answered their query.

WA’s chatbot Anna is based on ML technology. The chatbot only answers questions from citizens about rules and benefits and does not make decisions about actual cases. Anna’s analysis of the citizens’ requests is carried out in two rounds. First, the chatbot system will pre-process the citizens’ text by transforming it into an internal representation. This transformation process consists of several steps, including synonym replacement, spell checking, stemming, and removal of unknown words and stop-words. In this simplification process, stop-words such as ‘please’ and ‘if’ are removed, misspelled words are corrected, and words such as ‘remember’ and ‘pin code’ are reduced to a base root, i.e., ‘rememb’ and ‘pincod’. This step is performed before the citizen input is analysed by the ML model.

Secondly, the ML model will (try to) predict the citizens’ intent, i.e., classify the requests as belonging to a category of citizen questions. All questions in a category have the same intent and will give the same reply from Anna. An intent can be described as

a representation of the citizen request; it represents the variety of citizen formulations that can be matched to it. In the chatbot framework, intents are organized in intent trees. The root node of a tree, the so-called root intent, is the most generic intent of the tree. More specific intents are found at the end nodes of the intent branches. The intent trees are designed by chatbot trainers (more about these below). The trainers also create the test questions that the intent classification model is trained and tested on. These data sets are often based on real citizen input and augmented with typical questions that the chatbot advisors have experienced themselves.

The decision on whether a match is found is based on a set of prediction rules given by the software company that delivers the chatbot platform. These prediction rules constitute the internal chatbot framework. The chatbot will make a prediction based on the part of a citizen's input text that is probably the most important and will be based on the text's keywords about welfare terminology. Usually, this process is done in three steps (see Fig. 2), where the message analysis stops at the last valid step.

There are four core prediction rules and a few minor rules that determine if a prediction step is valid or not. The core prediction rules are:

- if a citizen input matches the training data perfectly, that input is automatically predicted at 100% match
- each step must predict an intent at 45% or more to be valid, i.e., the data that Anna's analysis is based on must indicate at least 45% probability for a match with the citizen's input
- a step will not be valid if the predicted intent belongs to a different root intent branch than the last predicted intent
- the top two predicted intents must have a difference of at least 15% to make a step valid.

We can see Anna as an information retrieval model where the knowledge base is a set of questions matched to answers. The strength of an information retrieval model is that it ensures the quality of the answers (Caldarini, 2022). All of Anna's replies are manually designed to give correct information about the laws and regulations related to the citizen's request. The replies are formulated by WA advisors, who know what kind of replies or answers the citizens need. If the system finds the citizen's request similar enough to an intent, the predefined answer linked to this intent will be given as Anna's reply. Anna is not expected to respond to complicated questions, as the WA neither considers the technology nor the citizens to be 'good enough' yet.

Anna is trained by chatbot trainers. These trainers are WA employees who work both as advisors and with improving and expanding the chatbot's repertoire of question formulations. The chatbot trainers carry out the work of finding and preparing the data

to be added to the chatbot model (Muller et al., 2019; Parmiggiani, 2022). Training Anna consists of adding intents: creating new intents implies that more questions are classified, and more answers are added to the system as responses to the questions. When new intents are created, Anna can answer a greater variety of citizen questions. The trainers continuously add more keywords and synonyms to the chatbot. Training data is used to ‘teach’ the chatbot how to ‘understand’ different citizen inputs and make predictions. As the chatbot is based on ML technology, the results will change over time as the chatbot system is fed with more intent-answer pairs. The ML that takes place consists of automatically adjusting the probabilities for the intents. The probability of an intent to be predicted will vary based on the frequency rate of that specific intent.

The chatbot trainers constantly monitor Anna. To further improve the chatbot, they read through reports generated by Anna, review chat conversations and analyse feedback on the chatbot’s performance provided by WA advisors. Anna has improved when the chatbot is better at matching citizens’ requests to relevant and correct intents.

4.2 Anna behind the façade

Information on how Anna has analysed a request is available in the chatbot’s administration panel. As the panel presents an overview of the chatbot’s predictions, it provides insight into how the chatbot works. By investigating these predictions, it is possible to get insight into why the chatbot matched a specific response to a given citizen input.

Fig. 2 shows how the chatbot’s tree structure is presented in the chatbot’s administration panel. The figure also shows how the chatbot compares its predictions of the likelihood of a possible match between a citizen’s request and the intents defined in the system. If the system does not find a sufficiently similar predefined intent, a standard fallback message is given. An example of a fallback message may be “Please rephrase your question in simpler terms”. A citizen request can also be marked as Unknown, meaning that the chatbot cannot connect the request to an intent. This may happen if a request concerns a topic not yet added to the system or is out of Anna’s scope. If two inputs in a row are predicted to have the same intent, Anna will reply that the request may not have been understood correctly. The chatbot will never reply to the same intent twice in a row. This function is implemented to make sure that the conversation does not result in a loop where Anna continues with the same replies. If the citizen tries to rephrase a request, but the new request is too similar to the previous formulation of the citizen, Anna will generate a fallback message. However, if a citizen formulates other requests in between, Anna may reply to the same intent several times during a conversation. The citizen may for this reason receive repetitive replies from the chatbot.

Fig. 2 shows that the system in Step 1 has calculated that there is a 73.71% probability that the citizen's intent is related to one found in the chatbot's 'General questions' (Generelle spørsmål) tree. As this number is above the 45% threshold, and at least 15% higher than the alternative 'About WA' (Om WA) tree, which is listed second in Step 1, the analysis continues to Step 2. In step 2, the intent 'Are you' (Er du) is calculated to 58.41% and the second alternative 'Talk to Human' (Snakke med menneske) is calculated to 14.30%. According to the system's rules, the analysis will then continue to Step 3. At Step 3 the intent 'Are you real' (Er du ekte) is calculated to 51.94%. This step is valid as it fulfils the rules given by the chatbot platform company, hence, Anna will reply with the answer corresponding to this intent. The root intent branch of 'Are you real' is 'General questions', and as an input will be analysed in accordance with the root intent branch of the last predicted intent, the next citizen input will be analysed in accordance with this intent. In the chat log presented in Fig. 2, the predicted intent is not correct. The citizen requests to be connected to a human, and hence, the Step 2 intent 'Talk to Human' is more likely to have given a more suitable response.

5 Research method

Our study is carried out as a qualitative interpretive case study (Stake, 1995; Walsham, 2002) where the methods for data collection were document analysis of logs and other documents, interviews, and participant observation. The study is documented in more detail in (Simonsen, 2019; Steinstø, 2020).

Message analysis: ?



kan du tilkoble meg til ekte mennesker, vær så snill? Norwegian Regular/Valid Answer Positive: 73%

Translated w/synonyms		Processed	
can you tilkobl myself to real peopl pleas		can you <unknown> myself to real peopl pleas	
Step 1	Step 2	Step 3	
1. X Generelle spørsmål 73.71%	1. Er du 58.41%	1. Er du ekte 51.94%	
2. X Om WA 14.96%	2. Snakke med menneske 14.30%	2. Snakke med menneske 14.30%	
3. X Foreldrepenger - IKKE SLETT 5.78%	3. Jeg (Kunde/Bruker) 6.28%	3. Er du bedre enn et menneske? 6.05%	

Figure 2. The three steps of the message analysis with probabilities for possible intents can be found in the chatbot's administration panel. The question from the user is translated into a kind of truncated English

5.1 Data collection

Chat logs: The data corpus for this study is a segment of chat logs retrieved from chat conversations between citizens and the WA's chatbot Anna from more than 6 months in 2019, each week including approximately 3000 chats. For this study, our data set is approximately 8000 unique chat logs we extracted from two weeks in May and two weeks in September 2019, which two of the authors browsed through and inductively familiarized themselves with. After the interviews with and observation of the chatbot trainers' work, where we learned about problematic chat situations, we carried out a more focused extraction of interesting logs. In the logs, the citizens are anonymous; hence, we have no information about them beyond what they write in the conversation with Anna.

Interviews: We interviewed advisors and chatbot trainers at two WA units. We carried out four semi-structured interviews with three different advisors, and three semi-structured interviews with two chatbot trainers who worked with improving Anna. We have applied what we learned in the interviews in the analysis of the relation between the citizens' requests and the chatbot's replies.

Participant observation: We carried out participant observation of advisors as they chatted with the citizens. The advisors explained how they interpreted what the citizen wanted to know from the citizen's question. They told us about typical issues that would imply that a chat with Anna gets transferred to the advisors. By interviewing and observing the advisors, we gained an understanding of common questions from the citizens and how the advisors usually would respond. They helped us to understand the relation between the citizens's requests and the chatbot's responses. By combining interviews and observations we acquired a deeper understanding of typical questions from the citizens, how the advisors understood the requests, and the responses from the advisors. This understanding was necessary for the chat log analysis.

We also observed chatbot trainers at two WA units while they trained the chatbot. They explained how the chatbot functioned and what they did to improve its responses. This helped us better understand how the chatbot works and where difficulties normally arise. The chatbot trainers guided us in finding and understanding the data provided by the chatbot administrative panel. These observations were useful for the analysis as we learned how the ML functioned.

Documents: We studied WA's plans and internal reports as well as the chatbot trainers' documentation of the chatbot framework and its functionality. The technical information was valuable for our analysis. Fully understanding the technical documentation required technical knowledge.

5.2 Analysis

The six chat logs presented in this paper are a result of three selection rounds. The selection process was aimed at documenting chat sessions where the response did not match the information needs that are expressed in the citizens' requests.

In the first selection round, we identified chat sessions where the citizen made explicit that they did not receive a good or relevant reply from Anna. In these chats the citizens said that the information is useless, they swear etc. or the conversation ends in an abrupt way. Some citizens also end the session by asking to chat with a human advisor. In the second round of selection, we identified chat sessions where there was no visible cue that the chatbot answer did not match the request and the chat session seemed to end in a good way. In these chats we looked for mismatches related to previously identified problems with language, citizens' understanding of chatbots, and their knowledge of welfare issues (Simonsen et al., 2020). In the third round we studied the ML-analysis of the inner workings to find a rationale for failed conversations.

The chatlogs were analysed with Suchman's analytical framework (1987). The framework helped us match the response from the chatbot with the citizen's requests, as it allowed us to focus on information given by the citizen that was not reflected in Anna's reply. What the citizen writes in the chatbot's input field is 'actions available to the machine' which is input to the interpretation of the ML language analysis. Anna's response in the chat is 'effects available to the user' in the analysis.

The framework also enabled us to include the citizen's 'actions not available to the machine' by interpreting the citizens' motivation for a question. For this analysis, we consider information about the citizen's life situation, which the citizen states but that gets lost in the processing, and therefore becomes 'not available to the machine'. Suchman's framework allows us to include the citizen's life circumstances as formulated by the citizen in our analysis and match these with Anna's response to evaluate the relevance of the chatbot's response. Information about the citizen's life situation will in many welfare related cases be crucial for giving a good answer, but the citizen is not always aware of which information is relevant.

In this study, we take the technology into account and describe how the ML carries out its calculations of the input text from the citizen, and discuss how this particular ML analysis has consequences for how the input from the citizen is interpreted and answered by Anna. We have used knowledge about the way Anna's predictions work to fill in column IV in the framework. We have found it illuminating to see the computations performed to produce an answer side-by-side with the citizens' requests and the answers by Anna. The framework enabled us to distinguish between the part of Anna's actions

that came from the citizens' formulations and those coming from the computer's inner workings.

To go into more detail about how the information given by the citizen was matched in the response from Anna, we applied the four levels of information needs suggested by Taylor (1968) to a selected set of chat logs. The four levels enabled us to analyse Anna's responses to what we interpreted that the citizen wanted to know from the start to the end of the chat session. We also wanted to identify how the chatbot contributed to improving the citizens' questions to get a better response. Our knowledge about what would be a correct and relevant reply to the original question was important for this analysis. In the next section we present six chatlogs where our analysis illustrates how mismatches occur.

The chats we have chosen for this article illustrate some problematic issues with the chatbot. The selection and narrowing down of chatbot logs for the analysis is not aimed at being a representative selection of chatbot logs. They act as illustrating examples that can show how the chatbot operates and why the functionality of the chatbot's inner workings can lead to unsatisfactory responses.

6 Into the chats

From the collection of chat logs, we have selected six logs that we present in more detail. An overview of the selected chat logs is given in Table 1.

<i>Log no</i>	<i>Log title</i>	<i>Date</i>
1	Going to have a child, what do I have to apply for?	May 10, 2019
2	Can I get money Anna?	Sept 13, 2019
3	How do I fill out an employment status form when I have a father quota?	Sept 13, 2019
4	Pregnant student, what am I entitled to?	May 15, 2019
5	Can you receive parental benefits and cash-for-care benefits at the same time?	Sept 15, 2019
6	Young disabled person married to a foreign single mother	May 14, 2019

Table 1. An overview of the chat logs.

Each chat log will be presented in a table with thorough descriptions of what is happening, both on the system's surface and of the mechanisms that play out in the background. The tables build on and expand Suchman's (1987) framework for human-machine interaction. In addition, we comment how Taylor's (1968) levels of information needs can characterize the citizen's requests. Each table shows an entire chat session between the chatbot and a citizen. The first column shows the citizen input, the next two columns show the chatbot's inner mechanisms and the fourth column shows the chatbot's response. The last column contains our interpretation of the citizen input and an explanation of Anna's reply.

We have labelled the buttons Anna presents in some replies with '(button)' to differentiate from regular bullet points in the text. When a button is pressed, the button is underlined. When the citizen clicks a button there will be no prediction of intent and we have made these places in the chat sequence grey, along with other areas of inaction. The column that shows the chatbot's mechanisms is divided in two, showing 'internal representation' and 'predicted intent'. This information is retrieved from the chatbot system and is not available to the citizen.

Log 1: Going to have a child, what do I have to apply for?

The citizen in this log expresses an information need and makes a first formulation of the need (Table 2). The question formulations miss keywords that can specify the request for information, and we have categorized the question formulations as level Q2 and Q3.

To the citizen's first question, Anna replies with a fallback message. This can happen for several reasons, e.g., that the system predicts two intents with approximately equal probability rates. For the citizen's second question, Anna manages to predict an intent. However, the chatbot's response about payments does not answer the original question and is probably not relevant for the citizen. Our interpretation of this log is that the citizen's information need concerns what kind of services and benefits an expectant parent may apply for. Anna does not help the citizen to formulate a question on the Q4-level and the citizen leaves the chat without receiving a relevant response.

Log 2: Can I get money Anna?

The citizen in this chat log (shown in Table 3) does not make clear what kind of money s/he needs. This citizen does not use welfare terms and does not express knowledge

Verne et al.: How Can I Help You?

User	Chatbot		Comments and interpretation	
Actions available to the chatbot	Mechanisms			
	Internal representation	Root intent / predicted intent	The chatbot's response	
			<p>My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about.</p> <p>How can I help you?</p> <ul style="list-style-type: none"> • For English chat, click here (button) 	<p>Anna's standard opening phrase.</p>
<p><i>Going to have a child, what do I have to apply for?</i></p>	<p>shall have child what must myself apply if</p>	<p>Unknown</p>	<p>Sorry, but I do not understand what you are asking. I will understand you better if you write short questions and include which benefit it is about. Can you try to ask the question in a different way?</p> <ul style="list-style-type: none"> • Need more help (button) • No (button) 	<p>This user expresses a need for information and makes a first formulation of the need without keywords from the welfare domain. We interpret this request to be on level Q2. Anna cannot predict an intent for the question, and a fallback message is generated which encourages the user to make another attempt.</p>
<p><i>How do I get money when I am expecting a child?</i></p>	<p>how get myself cash when myself wait child</p>	<p>Payment / Payment date</p>	<p>The day on which the payment is transferred depends on the benefit in question. You can see the payments from WA on Your payments on WA's website, but they will not be displayed until they have been credited to your account.</p> <ul style="list-style-type: none"> • Overview payout dates (button) • Overview Payout (button) • Not received money (button) 	<p>The user reformulates the question, and Anna's ML manages to predict an intent for it. The response does, however, not answer the user's original question. We interpret this as Q3 as the user's request is reformulated but does not give a relevant answer.</p>

Table 2. In this chat, Anna does not help the citizen to formulate a question that will give a good and relevant answer.

about the different WA benefits. Anna presents several buttons, each representing a possibility for further information retrieval, to the citizen.

<i>User</i>	<i>Chatbot</i>		<i>Comments and interpretation</i>	
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>			
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
			<i>The chatbot's response</i>	
			My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about. How can I help you? • For English chat, click here (button)	Anna's opening phrase.
<i>Hi</i>	hi	General questions / Hi	Hi! How can I help you?	The user says 'hi', and Anna responds.
<i>Help!!</i>	help	General questions / Help	How can I help you?	The user expresses a need for help, but does not formulate a proper question. We therefore see this as Q2.
<i>Can I get money Anna?</i>	can myself get cash jame	General questions / Disbursement	Under Your payments you will find an overview of all disbursement from WA in the last 3 years. Log in to Your payments (on the WA's website), or read more on the links. If you do not find what you are looking for in Your payments, please call us for assistance, <phone number>. • <list of different options regarding payments>	The user asks a question that does not include any terms that indicate any welfare topic. We see this at Q2. Anna responds with general info about payments.
<i>Am I entitled to child money?</i>	have myself entitd child cash	Unknown	Sorry, but I do not understand what you are asking. I will understand you better if you write short questions and include which benefit it is about. Can you try to ask the question in a different way? • Need more help (button) • No (button)	This question does not contain any keyword; 'child money' is not a term used by the WA. After three tries, the user has still not formulated any request that gives a relevant answer. We interpret this as Q2.
<i>I have had a child, can I get money?</i>	myself have get child can myself get cash	Unknown	Sorry, but I do not understand what you are asking. I will understand you better if you write short questions and include which benefit it is about. Can you try to ask the question in a different way? • Need more help (button) • No (button)	The user rephrases the question but does still not use keywords that indicate a welfare topic. We interpret this as Q2. Anna responds with general info.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>			
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		<i>The chatbot's response</i>	<i>Comments and interpretation</i>
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
<i>I have had a child</i>	myself have get child	Parental benefit / I am or I am going to be a mother	Maybe some of these can be helpful? <ul style="list-style-type: none"> • Parental benefit (button) • Lump-sum grant (button) 	Now the chatbot matches the user statement 'I have had children' with the intent 'Parental benefit'. The chatbot replies with listing key benefits regarding having a child. We still interpret the request as Q2.
<i>Parental benefit</i>			The parental benefit is supposed to ensure the parents income in connection with birth and adoption. What do you want to know? <ul style="list-style-type: none"> • Am I entitled? (button) • <more options regarding parental benefit> 	User clicks the 'parental benefit' button and the chatbot responds with buttons that show alternative responses. By interacting with the chatbot through the use of buttons, the user can 'ask' precise questions that the chatbot can answer. We interpret clicking this button to be Q4.
<i>Am I entitled?</i>			The earning rules are the same, regardless of what occupational status (employee, self-employed, freelancer) you have. Benefits such as transitional benefits, Cash-for-care benefits and disability benefits do not provide earnings for parental benefits. Are you or will you be: <ul style="list-style-type: none"> • Mother (button) • Father (button) • <more options> 	User clicks the 'Am i entitled' button and the chatbot provides information regarding earning rules.
<i>Father</i>			As a father, you can earn the right to parental benefit until you start withdrawing parental benefit, and no later than the last day of the joint period. You may be entitled to parental benefit if you have worked and have had pensionable income for a total of at least 6 of the last 10 months before you start receiving parental benefit. Some WA benefits are equated with work. Do you have the following? <ul style="list-style-type: none"> • Yes (button) • <more options> 	User clicks the 'Father' button and receives information about parental benefits for fathers.

<i>User</i>	<i>Chatbot</i>		<i>Comments and interpretation</i>	
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>			
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
		<i>The chatbot's response</i>		
<i>What is a joint period?</i>	what are joint period	Parental benefit / What is a joint period?	The joint period is the weeks that remain when the 3 weeks before the term and the quotas of mother and father or co-mother have been deducted. How long the joint period will be depends on whether you have chosen 100 percent or 80 percent parental benefit, when the child is born or adopted, and how many children that are born or adopted. If the father or co-mother is to use the joint period, there are requirements for what the mother does. We call this activity requirements. <ul style="list-style-type: none"> • Activity requirements (button) • <more options> 	The user clicks the button for 'joint period' and receives a reply that we interpret as relevant to this question. We see this as Q4.
<i>Yes</i>			To meet the income requirement, your calculation basis must be at least <currency> 49,929 (½ G). If your parental benefit period started on 30.4.19 or earlier, your income during the accrual period converted to annual income must be at least <currency> 48,441 (½ G). Do you meet the income requirement? <ul style="list-style-type: none"> • Yes (button) • No (button) 	The user clicks the 'yes' button from earlier in the conversation.
<i>What does ½ g mean?</i>	what mean one two g	General questions / What is national insurance basic amount (G)	National Insurance scheme basic amount (often abbreviated to G) is used when determining pension points and calculating benefits from WA. The amount is adjusted every year. As of May 1 2019, the basic amount is (1 G) <currency> 99,858. Here are some common amounts based on the base amount: <ul style="list-style-type: none"> • ½ G = <currency> 49 929 • 3 G = <currency> 299 549 • 6 G = <currency> 599 148 • Overview on WA's website (link) 	The user asks about ½ G and receives an explanation for the term. In short, G is an abbreviation for the National Insurance Basic Amount, a standard amount that is used to calculate benefits and pensions and which is set on 1 May each year. 1 G = 1 x Basic Amount, 2 G = 2 x Basic Amount, etc.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>		<i>Comments and interpretation</i>	
	<i>Mechanisms</i>	<i>The chatbot's response</i>		
<i>Actions available to the chatbot</i>	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
<i>What is the basic amount?</i>	what are basic amount	Perfect Match (Double Hit)	Maybe I did not fully answer what you want to know? Maybe you can ask a little differently?	This question is concerning the basic amount. In Anna's last reply, an explanation of this term was given. The user's question might indicate that the user did not read this explanation, did not understand it or wanted a more thorough explanation. As Anna predicts the same intent twice in a row, the chatbot replies that it does not understand the question asked.
<i>When do I get money?</i>	when get myself cash	General questions / Payout date	The day on which the payment is transferred depends on the benefit in question. You can see the payments from WA on Your payments on WA's website, but they will not be displayed until they have been credited to your account. <ul style="list-style-type: none"> • Overview payout dates (button) • Overview Payout (button) • Not received money (button) 	As the question is formulated without the proper terms, we interpret this as Q3. The question is still answered by the chatbot.
<i>Thanks Anna</i>	thank jame	General questions / Thanks	You're welcome!	The user thanks the chatbot. This may indicate that the user is satisfied with the answers and/or wishes to end the conversation.
<i>Nice to taye (sic) with you</i>	great <unknown> with you	Possible Missing Intent	I am not sure if I have understood you correctly, but can I help you with: <ul style="list-style-type: none"> • How are you (button) 	The chatbot does not recognize the misspelled word, which might be the reason why the chatbot does not 'understand' what the user is writing. As the chatbot sometimes lists possible intents in order to limit the false positive ratio, some awkward situations may arise.
<i>How are you</i>			I am fine.	The user clicks the 'How are you' button.

Table 3. In this chat, Anna presents buttons that may help the citizen.

As the citizen's request are not formulated in the relevant terms for the topic, we see this as an expressed information need starting at level Q2. Anna presents buttons which

prompt the citizen to iteratively specify the request and to further questioning. By offering buttons that show possibilities for more precise information retrieval, with correct keywords and within Anna's scope, the citizen is helped to express requests at a higher Q-level. This citizen receives an answer to the original question. The "Thanks Anna" indicates that this citizen finds the answers satisfactory.

Log 3: How do I fill out an employment status form when I have a father quota?

In the chat log shown in Table 4, the citizen asks how to fill in the "employment status form", but this form is outside the chatbot's scope. However, the question concerns the father quota, which is within the scope. Anna does not make clear that unemployment is outside its (current) scope and does not direct the citizen to a different communication channel. The citizen is not helped to a higher Q-level and leaves the chat without receiving a relevant answer.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>			
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		<i>The chatbot's response</i>	<i>Comments and interpretation</i>
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
			<p>My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about. How can I help you?</p> <ul style="list-style-type: none"> • For English chat, click here (button) 	<p>Anna's opening phrase.</p>
<p><i>Hi! How do I fill out an employment status form when I have a father's quota?</i></p>	<p>hi how fill myself out employment status form when myself have dad leav</p>	<p>Parental benefit / father's quota</p>	<p>The father's quota is the part of the parental benefit period that is reserved for the father. It is often called paternity leave.</p> <ul style="list-style-type: none"> • Length of the quota (button) • Divide the father's quota (button) • Am I entitled? (button) • How do I apply? (button) • Transfer to mother (button) 	<p>The user asks a question that is out of the chatbot's scope. The chatbot is not yet trained to answer questions about employment status forms. Anna's response informs the user that 'paternity leave' is a term that can be used as a synonym for the term 'father's quota'. The question contains some keywords that indicate the welfare domain, but does not give an answer. We interpret the request to be Q3.</p>
<p><i>What am I supposed to write in the employment status form when I have paternity leave??</i></p>	<p>what shall myself write employment status form when myself have dad leav</p>	<p>Parental benefit / Application parental benefit</p>	<p>I am happy to help you with the parental benefit application. Remember that it is the person who is to have the parental benefits that sends the application. What do you want to know?</p> <ul style="list-style-type: none"> • How do I apply? (button) • When should I apply? (button) • Residence (button) • Changes (button) • Application status / case (button) • Attachment (button) 	<p>This question is a reformulation where among other things, the term 'paternity quota' is replaced with Anna's proposed term 'paternity leave'. The system predicts that the user's question is associated with the intent 'Application for parental benefit' and Anna's reply contains information about the parental benefit application. The user ends the conversation here, and we interpret the request to be Q2.</p>

Table 4. In this chat, the citizen asks a question involving unemployment status forms, which is outside of Anna's scope.

Log 4: Pregnant student, what am I entitled to?

In this chat, the citizen asks several questions that seem to be answered by the chatbot (Table 5). The sequence of requests is within the system's scope, contains the correct terms and looks like Q4 requests. Anna responds that the citizen may not be entitled to parental benefit and suggests that the citizen reads more about earning rules. The chatbot does not provide information about another benefit that may be relevant for this citizen, namely the lump-sum grant (see Log 2) available for students without an income. Although the internal representation recognizes that the citizen is a student, this is not present in the responses. That the citizen is a mother is recognised by the chatbot, but the relevant benefit is not mentioned in the answer.

<i>User</i>	<i>Chatbot</i>		<i>Comments and interpretation</i>	
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>			
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
		<i>The chatbot's response</i>		
			My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about. How can I help you? • For English chat, click here (button)	Anna's opening phrase.
<i>Pregnant student, what am I entitled to?</i>	pregnant student what have myself entitl	Parental benefit / Earning rules parental benefit	The earning rules are the same, regardless of occupational status (employee, self-employed, freelancer) you have. Benefits such as transitional benefits, cash benefits and disability benefits do not provide earnings for parental benefits. Are you or will you be: • <u>Mother (button)</u> • Father (button) • Co-mother(button) • Foster parents (button)	This user expresses an information need when pregnant, but does not use terms for WA's services and benefits. We interpret this request to be on level Q3. Anna predicts the root intent to be 'Parental benefit' and gives a generic answer for this topic.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>			
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		<i>The chatbot's response</i>	
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
<i>Mother</i>			<p>At birth, the parental benefit period starts no later than three weeks before term. Upon adoption, it starts on the date of taking over care. You may be entitled to parental benefit if you have worked and had a pensionable income for a total of at least 6 of the last 10 months before the parental benefit period begins. Some WA benefits are equated with work. Do you have this (click on what applies to you)?</p> <ul style="list-style-type: none"> • <list of other options> • I am a student (button) 	<p>Anna gives more information which is relevant for a coming mother-to-be and requests more specific info about the user. The user selects 'student'. The buttons present a selection of choices that can be relevant to the user as part of a question negotiation with Anna.</p>
<i>I am a student</i>			<p>The rules for earning parental benefits are the same for everyone, including students. Have you had pensionable income 6 of the last 10 months? Loans and scholarships from the National Educational Loan Fund are not considered pensionable income.</p> <ul style="list-style-type: none"> • Yes (button) • No (button) 	<p>Anna presents information about what applies to students who are to become parents. As the user receives information of high relevance to the original question, we interpret this request to be level Q4. Anna asks for specific information about the user's income to give a more precise answer.</p>
<i>No</i>			<p>Then it is not certain that you are entitled to parental benefits. Read more about the earning rules on WA's webpage</p> <ul style="list-style-type: none"> • Earnings on WA's webpage (link) 	<p>As this user clicks 'no' to having had pensionable income 6 of the last 10 months, she is informed by Anna that she may not be entitled to parental benefits. The user clicks a relevant button, and hence, we interpret this as Q4. Anna's answer is formulated vaguely by the chatbot trainers as they do not want Anna to give precise answers about what a user is and is not entitled to.</p>

<i>User</i>	<i>Chatbot</i>		<i>Comments and interpretation</i>
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>	
<i>Ok</i>	ok	General questions / Ok	Is there anything else I can help you with?
			By answering 'Ok', the user is apparently accepting Anna's answer. The user does, however, not receive information about the lump-sum grant that may be relevant for a student who is becoming a mother.

Table 5. In this chat, Anna does not inform the citizen that she may be entitled to another benefit, the lump-sum grant. The citizen cannot see from the answer that this information is missing unless she already knows about the lump-sum grant.

When Anna states that the citizen is not entitled to parental benefit, the citizen replies 'ok' instead of asking any further questions, hence, the citizen does not receive a complete answer to the original question "what am I entitled to". We categorize this question as a *disguised Q4*: the questions seem to be asked in a correct way and receives a relevant answer, but the answer is not complete. A disguised Q4 cannot be identified from the text only; to detect that important information is missing in the answer requires knowledge about alternative benefits that may be relevant.

Log 5: Can you receive parental benefits and cash-for-care benefits at the same time?

In the chat log in Table 6, the first question from the citizen mentions two different WA benefits, and Anna is not able to predict one intent for the question. The question is well formulated and contains the correct terms. The citizen is presented with a sequence of buttons that in the end gives an answer to the first question. The citizen must follow the chatbot's internal structure in order to receive the information that is requested. This is an example of how the structure of the intent trees affect the conversation flow.

One possible explanation for this example is that the chatbot does not (yet) contain an intent for this question. Another explanation is that the chatbot cannot handle two keywords in the same request. Even though the chatbot does not predict an intent, it

User	Verne et al.: How Can I Help You? <i>Chatbot</i>			
Actions available to the chatbot	Mechanisms		The chatbot's response	Comments and interpretation
	Internal representation	Root intent / predicted intent		
			My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about. How can I help you? • For English chat, click here (button)	Anna's opening phrase.
<i>Can you receive parental benefits and cash-for-care benefits at the same time?</i>	can someone get parent benefit and cash benefit simultan	Possible Missing Intent	I might not have understood everything you asked, but I can help you with this: • Cash-for-care benefit (button)	The user's question mentions two benefits which both are recognised by the chatbot. Anna only responds with the 'cash-for-care' benefit. The question contains the proper benefit terms but there is no intent containing both. We interpret the question to be Q3 as it is not adapted to the information system used.
<i>Cash-for-care benefit</i>			Cash-for-care benefit is a tax-free benefit that can be given for children between 1 and 2 years of age who do not attend a full time government subsidized kindergarten. A part-time place in kindergarten can give the right to partial cash-for-care benefit. The person who applies must have at least 5 years of insurance in Norway and/or EEA countries to be entitled to the cash-for-care benefit. If the child lives with both parents, both parents must fill the condition of 5 years of insurance. • About application and how to apply (button) • Stop receiving cash-for-care benefit (button) • How long can I receive cash-for-care benefits? (button) • Relation to parental benefit (button) • How much will I receive? (button)	Anna informs the user about cash-for care, and the answer contains buttons for further information. The user clicks the 'Relation to parental benefit' button. Apparently, clicking the cash-for-care button did not give a complete answer but provided buttons that helped the user in the right direction. We categorize the request as Q3 as it contains the correct terms but does not meet the information needs.
<i>Relation to parental benefit</i>			If you receive parental benefit at birth, you may receive parental benefit and cash-for-care benefit at the same time. Neither parental benefit nor cash-for-care benefit will be reduced if you receive both. Different rules apply for parental benefits in relation to adoption. Call us for more information, call 55 55 33 33 and press the button for Family.	Anna replies with information about the relation between parental benefit and cash-for-care, which was what the users initially asked for. As this button gives the user a complete and correct answer to the initial question, we interpret the request as Q4. The user receives an answer to the initial question, but has to use buttons to get there.

Table 6. This chat log illustrates how a citizen has to follow the internal structure of the chatbot in order to receive the information needed.

does suggest one possible intent, and thereby helps the citizen in the right direction. To be able to access the intent branch 'Relation to parental benefit', the citizen has first to access the tree for 'Cash-for-care'. Anna cannot answer a well-formed question when its analysis is in the wrong branch of the intent structure. The level 3 question was helpful for directing the conversation to the correct topic (branch) before asking the level 4 question. After question negotiation in a sequence of clicking buttons, the citizen receives an answer that meets the information needs expressed in the first question.

Log 6: Young disabled person married to a foreign single mother

In the chat log shown in Table 7, the citizen's first question contains an inconsistency: if you are married you cannot be a single mother. The citizen clicks on the buttons Anna provides, but the answers do not fully match the situation the citizen describes as they all focus on benefits for a single mother. The answers are all 'correct', i.e., the information provided is written by skilled WA advisors. However, the information is not relevant for the circumstances stated by the citizen as information about benefits for single mothers is of no use if the mother is married.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>			
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		<i>The chatbot's response</i>	<i>Comments and interpretation</i>
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
			<p>My name is Anna and I am a chatbot. I am best adapted to answering short questions. Remember to include what type of benefit you are asking about. How can I help you?</p> <ul style="list-style-type: none"> • For English chat, click here (button) 	<p>Anna's opening phrase.</p>
<p><i>can a young disabled person married to a foreign single mother receive child benefit and child supplement</i></p>	<p>can child disabl are <unknown> with abroad sole caregiv get child benefit and child add</p>	<p>Child benefit / Extended child benefit</p>	<p>If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children. The extended child benefit is a child benefit for one more child than what you actually are living together with. You have to apply for an extended child benefit, it will not be automatically granted.</p> <ul style="list-style-type: none"> • <u>Who is entitled to extended child benefit? (button)</u> • <list of other options regarding extended benefits> 	<p>This question contains many keywords. Anna recognizes one topic (extended child benefit) and does not recognize the contradictory words 'single' and 'married'. Anna's answer is not relevant as it is only applicable for someone who is not married. We see this question as Q3 as it is not adapted to the information system Anna by using too many keywords.</p>
<p><i>Who is entitled to extended child benefit?</i></p>			<p>If you are a single mother or father you can receive extended benefit if you are living alone with the child and:</p> <ul style="list-style-type: none"> • you are separated, divorced or widowed • <continues to list criteria for extended benefit> <p>If you do not have common children with the person you have been living together with, we do only need documentation that you are not living together. It could be:</p> <ul style="list-style-type: none"> • Declaration of breakup • Separation permit • <u>When does the right to extended child benefit cease? (button)</u> • Apply for extended child benefit (button) • Shared child benefit (button) 	<p>User clicks a button provided by the chatbot about who is entitled to extended child benefit and receives information about the benefits for single mothers and fathers. We categorize this as a disguised Q4 as the answer does not match the situation previously described by the user.</p>

<i>User</i>	<i>Chatbot</i>		<i>The chatbot's response</i>	<i>Comments and interpretation</i>
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>			
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
<i>When does the right to extended child benefit cease?</i>			<p>Extended child benefit will cease if:</p> <ul style="list-style-type: none"> • 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 	The user clicks a button regarding the rights to extended child benefit. Anna informs the user that the right to extended benefit will cease if one marries.
<i>can a foreign single mother receive child benefit and child supplement</i>	can abroad sole caregiv get child benefit and child add	Child benefit / Extended child benefit	<p>If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children. The extended child benefit is a child benefit for one more child than what you actually are living together with. You have to apply for an extended child benefit, it will not be automatically granted.</p> <ul style="list-style-type: none"> • Who is entitled to extended child benefit? (button) • <list of other options> • Payment (button) 	The user asks if a foreign single mother can receive both child benefit and child supplement. Anna provides general information on extended benefits for a single mother or father.
<i>Payment</i>			<p>Child benefit will be paid out on the last day of the week each month. Child benefit is not liable to tax. Remember that you can see your payments from WA by logging into Your WA. At WA's website you can find an overview of payment dates for child benefit.</p> <p><lists payment info></p>	The user clicks the button 'payment' and receives general information regarding dates for child benefit payment.

Verne et al.: How Can I Help You?

<i>User</i>	<i>Chatbot</i>			
<i>Actions available to the chatbot</i>	<i>Mechanisms</i>		<i>The chatbot's response</i>	<i>Comments and interpretation</i>
	<i>Internal representation</i>	<i>Root intent / predicted intent</i>		
<i>Can a foreign single mother receive child supplement</i>	can abroad sole caregiv get child add child benefit	Parental benefit / Only mom is entitled	<p>In some cases, the mother may receive the entire period of child benefit. She can be alone with the care or live together with the child's other parent that has not yet served up the right to child benefit.</p> <ul style="list-style-type: none"> • I am caring for the child alone (button) • I live together with the father/co-mother (button) • <u>I mean child benefit (button)</u> • I mean transitional benefit (button) 	The user continues to ask about what a foreign single mother can receive and Anna responds with details about single mothers. The user seems to focus on the mother being foreign, and does not pick up that a single mother is not married. We categorize the question as a disguised Q4.
<i>I mean child benefit</i>			<p>If you are a single mother or father, you may be entitled to extended child benefit if you are living alone with children. The extended child benefit is a child benefit for one more child than what you actually are living together with. You have to apply for an extended child benefit, it will not be automatically granted.</p> <ul style="list-style-type: none"> • Who is entitled to extended child benefit? (button) • <list of other options> 	The user clicks the 'I mean child benefit' button and receives the same information as earlier in the conversation. The chatbot is designed to not provide the user with the same reply twice in a row. It may, however, give the same reply to the same intent several times during a conversation.

Table 7. In this chat, Anna does not detect that the citizen's question includes an inconsistency.

The inconsistency in the initial question is not commented on during the chat session by neither Anna nor the citizen, although Anna answers that a single mother has to live alone with the child (answer no. 3) and that extended benefit will cease if you marry (answer no. 4). However, there is no sign that the citizen has read this information as the citizen continues asking about extended benefits. We can only speculate if the citizen has read and understood the text in Anna's replies and wants to explore the single mother benefits or if s/he has not read or understood that a mother who is married is not seen as single. Anyway, Anna does not detect or correct the inconsistency in the initial question during the chat session, and the chatbot follows up with information that can mislead an ignorant citizen into believing that the married mother is entitled

to benefits for single mothers. The citizen's sequence of questions about benefits for a single mother are all categorized as disguised Q4 as each contains relevant keywords and receive a correct answer, but where important information provided earlier (about being married) is not taken into account. The citizen needs to know who counts as single mothers to find out that this information is not relevant for a married mother, as it is not explicitly mentioned in the answer.

6.1 Summing up

Our analysis focuses on the match between the citizens' requests and the chatbots answers, and if and how the chatbot helped the citizens formulate a request that give a better match, i.e., a better answer. In some conversations, the chatbot responds with irrelevant or misleading information to the citizens' questions. The logs show how the chatbot does not always meet citizens' information needs:

- Log 1: The citizen does not receive help to formulate a better request and does not get a good answer. The citizen seems to see that the answer is irrelevant.
- Log 2: This citizen receives help from the chatbot to get a relevant answer, which answers the original request and also the follow-up questions.
- Log 3: The citizen ends the chat as the answers are not becoming more relevant. This citizen does not receive help to improve the question but recognizes that the answers are unsatisfactory.
- Log 4: This citizen receives help in formulating better questions and it seems that the answer is relevant. However, the citizen cannot see from the final answer that there might be yet another benefit that can be relevant. We categorise the question to be a disguised level 4 question within this chatbot's model structure.
- Log 5: The citizen does not immediately receive an answer to the first question, but the chatbot helps by presenting buttons so that the citizen finds a relevant answer. The citizen has to patiently follow the chatbot's internal intent structure to get an answer.
- Log 6: The citizen asks an inconsistent question which is not recognised as such by the chatbot, which provides a sequence of responses about one of the welfare topics involved (single mother). This is another occurrence of disguised level 4 questions in that the citizen may not understand from the answers only that important information is missing in the response. Earlier research has confirmed that citizens do not read all the text in Anna's answers, especially not if the beginning of the text is automatically scrolled out of the communication window (Simonsen, 2019; Steinstø, 2020).

Citizens who have not stated their information needs with the right keywords, i.e., they express a need for information on level Q2, will experience problems receiving relevant answers from the chatbot. Many citizens write long questions as if they talk to a human when chatting with Anna. Long questions often contain several keywords, increasing the chance for the chatbot to predict a wrong intent, i.e., not finding a good match between the question and an answer. A question where the citizen uses the correct and relevant keywords (level Q4) has a better chance of being matched with the correct intent that answers the question. Anna may omit information if the citizen does not ask for it directly, like the pregnant student in Log 4 who did not get the information that she may be entitled to a lump-sum grant.

The chatbot's analysis of the disguised level 4 questions provides the most problematic examples in our data as the citizens receive a seemingly correct and relevant response but where important information is missing. The answer is correct but does not match these citizens' explicitly stated life situation: the chatbot answers are general and may or may not include information necessary for assessing if the answer is relevant to the situation at hand. In the example above (Log 6), we cannot see any sign that the citizen understood that questions about single mother's benefits are not relevant if the mother is married. The chatbot's responses did not invite the citizens to ask for clarifications or more information during these chat sessions. The citizens may find out later that they have not received enough relevant information. However, these questions are only *disguised as correct* relative to the information system of this particular chatbot at the time of the analysis because the intent-answer pairs that make up the knowledge base will change over time when the chatbot is trained. The inner workings of Anna at the time when the chat takes place frames which questions can receive relevant and sufficient answers.

7 Discussion: A chatbot for information needs

Our analysis of chat logs has identified some problems that can occur in conversations between citizens and chatbots, and why they occur. These problems become more serious when the chatbot Anna is the first or only representative of the WA that the citizens meet when they seek information and advice.

7.1 Matching questions to answers

The chat logs we have presented demonstrate that well formulated questions to the WA do not come into the citizens' mind fully formed. Using Taylor's (1968) four levels of

information needs to analyse the logs has helped us see how the chatbot—as an information system—analyses the questions to match a question to an answer. The process of matching shapes the answerable question.

In her study of phone calls to tax authorities, Verne (2015) argues that system understanding is necessary for the callers to be able to ask good and relevant questions. Our study of citizens' chatbot communication with a public agency confirms this: posing questions that receive good answers is correlated with having domain knowledge (Simonsen et al., 2020; Skaarup, 2021). Domain knowledge in this case includes both knowledge about the welfare field and how to adjust the question to the chatbot in a way that gives a correct and relevant response. People who do not themselves have relevant domain knowledge can benefit from negotiating their question with a specialist in the domain area. Domain knowledge helps with formulating questions that express the information needs precisely. Using relevant keywords correctly in the chatbot requests, expresses a kind of domain knowledge.

Our data shows that citizens who ask questions that the chatbot cannot answer receive one of two types of help to improve their question: they are asked to reformulate their question, or they can choose between a limited number of buttons. Reformulating the question in simpler words may lead to keywords that better match Anna's intents and therefore its answers. Clicking a button bypasses the ML analysis of the question and gives a direct pre-defined answer from the chatbot. We can interpret the presentation of buttons to click as Anna's contribution to question negotiating with the citizens (helping them raise the Q-level). The buttons can be seen as correctly formulated or 'compromised' questions that have answers. In this sense, they could be categorized as level four questions. However, the buttons and the information they point to may not be relevant for the citizen and we therefore often see them as level three questions. Level four is reserved for a good match between the request and the answer. In contrast to human advisors who teach callers the relevant key terms for their request or ask follow-up questions to give a good reply, the chatbot have quite limited means of helping a citizen to formulate answerable questions. Moreover, the route to a satisfactory answer for the citizen through clicking buttons can be rather cumbersome.

In our analysis we have categorized a well-formed question where a relevant keyword is used to be on level four (Q4), indicating that the question can be answered satisfactory by Anna. However, Anna interpreted some of the well-formed questions in a way that did not give a complete answer; we called these *disguised level four questions*. Our analysis shows that these questions were formulated with the correct keywords for the topic (as a level four question), but that Anna did not respond with a satisfactory answer. We want to emphasize that a seemingly correct response can be

problematic: where the chatbot does not address a misunderstanding, e.g., ‘married to a single mother’ or fails to inform about a benefit, e.g., the lump-sum benefit. The answers from Anna look correct, but our interpretation from the citizens’ description of their life situation in some of the chats is that the answers do not meet their information needs. Moreover, the citizens cannot assess whether the answer is satisfactory from the chatbot’s replies alone. This adds weight to the discussion about citizens’ domain knowledge above; one needs domain knowledge to ask good questions, but even more domain knowledge to assess if the answer is satisfactory.

In such cases the chatbot Anna does not fill its role alone as it is not able to interpret all questions that are answerable within the welfare domain. In contrast, human advisors often provide information that is not asked for when this is relevant for what the citizens express. Verne’s (2015) study of calls to the tax administration showed that the human call advisors adapted their response to (their understanding of) the callers: what the callers said about their life situation and circumstances, in particular if a life crisis triggered the call. Many of the callers had first searched the Internet and called the tax agency to confirm their understanding of how tax rules and regulations applied to their own life situation. The call advisors asked questions that clarified the issues so that the callers were able to follow up their tax issues on their own (Bratteteig and Verne, 2012). Callers who used incorrect keywords or otherwise expressed little knowledge about tax, received more help from the call advisor (Verne 2014; 2015). Bratteteig and Verne (2012) argue that it is important for the citizens in a democracy to have knowledge for questioning the government’s rules and regulations, i.e., have insight into governmental decisions and sufficient knowledge to be able to appeal a decision. Explanations about decisions will enable citizens without domain knowledge to question the governmental decisions (Amersi et al., 2019). Receiving explanations for advice about benefits will be important for the citizen because explanations contribute to the transparency of the chatbot’s decision. Transparency of decisions is one of the main topics in AI ethics guidelines (Dignum, 2018; Jobin et al., 2019).

Our study of chat logs also demonstrate that information needs are often complex, and it is difficult to know if the information need is met. Here we refer to Taylor’s original interview study of library visitors’ dialogues with librarians to find literature that met their information needs (1968). Taylor’s method could not assess how satisfied the library visitors were with the suggested literature after leaving the library. Our study has a similar limitation as we have not interviewed citizens using the chatbot. It would have been interesting to learn about the users’ opinions about if and how their information needs were met. As we did not have access to interview citizens in this study, we have discussed our interpretations of the questions with advisors and chatbot trainers as a

basis for understanding the citizens' questions and the chatbot's responses. We have been careful not to make assumptions about the citizens beyond what we can gather from the logs.

7.2 The hidden parts of interaction with chatbots

The inner workings of the chatbot plays a role in how Anna responds. In our analysis we modified Suchman's (1987, 2007) framework for human-machine interaction adding the mechanisms and calculations of the chatbot in the analysis. The correctness and relevance of the chatbot responses are heavily influenced by the hidden parts of the chatbot—which its users do not know: The input pre-processing, the vocabulary, the internal representation, the prediction rules, and intent structures. The main issue here is that the interpretation of the words in a citizen's input sentence will be analysed and assigned the intent with the highest probability. If the question produces more than one potential intent or if the processing of the question does not produce any obvious intent (e.g., if the citizen uses the wrong terms), the chatbot cannot select an intent and therefore cannot give an answer. The intent analysis can result in different intent trees for questions that are formulated differently but express the same information needs, leading to different responses from the chatbot to the same question depending on how it was formulated. Training the chatbot with more keywords and intents will not necessarily help as the ML-based inner workings are based on probabilities. The inner workings of a ML-based system that process input to output cannot today guarantee fairness (Teodorescu et al., 2021). This also holds for Anna's matching of questions to answers.

The ML-based analysis of the citizen's input can lead to misinterpretations, as the layers and steps of the analysis are hidden from the citizen. Sometimes the citizen had to navigate Anna's somewhat rigid hidden (tree) structure in search of answers. When the citizen receives an answer from Anna that does not answer the citizen's questions, the citizen does not know what went wrong or how to rephrase the question. The inner workings of the chatbot Anna have their own opacity (Burrell, 2016). We have seen examples in our data where Anna's spell checking wrongly corrects a word, a synonym is wrongly replaced, or the citizen input is connected to the wrong intent. In these cases, the citizen only sees the wrong answer from Anna, without knowing where it went wrong. As is often the case with AI and ML, small pieces of input data can make a huge difference in the interpretation. Changing one single pixel in an image can radically change how the image is classified (Antun et al., 2020). Similarly, a single word can affect the sentence prediction; if the citizen simply had removed one word, Anna could have given a different answer. Citizens cannot be expected to have knowledge about the

technical preconditions for the chatbot system's interpretations and answers. Our study shows that the chatbot users formulate their requests in line with Taylor's description of how the library users behave: they do not ask a well-formed question but give "a description of an area of doubt in which the question is open-ended, negotiable, and dynamic" (1968, p.179).

A precondition in Taylor's study is that the information to be found exists as files and books in the library shelves accessed through a paper-based file system. A good question on level four takes into account what can be found and how to find it, in Taylor's words: "At the fourth level the question is recast in anticipation of what the files can deliver" (1968, p.182). This description is also valid for Anna, rephrased as: at level four the question needs to be formulated in anticipation of the intents that can be matched with answers and how the chatbot's ML interprets the request. In theory, all of Anna's buttons therefore suggest potential level four questions as they all lead to an answer. In practice, however, buttons that are not relevant to the citizen's question are part of the question negotiation, and we therefore have categorized them as level 3.

7.3 Responsibility for correct and relevant information

WA is responsible for giving correct and relevant information about welfare rules and regulations to the citizens. The chatbot Anna is presented as the first point of contact on their website and citizens are led to begin there. The WA cannot expect citizens to have any knowledge about the chatbot or its inner workings nor how they can improve their questions by negotiating with these inner workings. Welfare clients are often in life-changing circumstances and have their minds there. In a study of information needs in healthcare, Timmins (2006) found that an information need arises in a person as a response to major life changes, external events or, e.g., incapacitating illness—a description that also fits information needs about welfare. Information needs occur as part of the individual's efforts to cope with the situation. These needs are unique to an individual and subjective in nature (Timmins, 2006) and speaking with a human advisor can help the citizen to match the rules with the individual circumstances. Tyckoson (2015) argues that the need for a domain expert increase if the information is structured according to the domain or the domain organization's processing of the domain, like rules in different welfare domains or organizational units. When talking with an advisor, the citizen can learn more about the welfare domain. This will help the citizens to receive good answers next time they ask a chatbot.

We argue that chatbots that inform about governmental rules and regulations have success criteria that go beyond chatbots' social interaction characteristics (Chaves and

Gerosa, 2021); it is more important that they present information that is correct and relevant for the citizens. Most citizens want to behave according to the law and feel “respected and recognized as a valued member of society” (Skaarup, 2021, p. 90) also when they receive welfare benefits. The Scandinavian countries have strong traditions for a fair distribution of their public services to those in need, including health and welfare services. If the citizens do not get sufficiently correct and relevant information, or misunderstand which rules apply to their own life situation, they risk becoming vulnerable: missing out on benefits or having to pay back benefit money that they have erroneously received. This is one reason that the disguised level four questions are problematic; the chatbot trainers we interviewed confirmed that the worst-case scenarios are situations where the citizens perceive Anna’s answers to be valid, when they are not.

Giving correct and relevant answers to the citizens’ information needs is a form of responsibility (Floridi et al., 2018). We argue that the ambition to give good answers, i.e., correct and relevant information needs to be evaluated as a response to a question expressing an information need. Giving good answers therefore needs to involve support of the formulation of ‘good questions’, i.e., questions that can get good answers. Negotiating a question to a well-formed and precise expression of a citizen’s information needs (level four) can be seen as an important part of the responsibility for giving correct and relevant answers.

In this paper, we have shown how a ML-based chatbot interprets citizens’ questions, and that this interpretation is what decides if the questions are answerable or well-formulated. The citizens cannot see how their questions are interpreted. With the exception of the buttons that show alternative possible intents, there is little transparency for the citizen about the ML’s interpretations or predictions that process and shape answerable questions². Transparency is a requirement for responsible AI (Amershi et al., 2019; Jobin et al., 2019). Making the inner workings of the chatbot’s predictions—or some elements of it—visible can improve the chatbot’s accountability towards the citizens (Amershi et al., 2019; Shollo, 2020). Technological transparency and accountability are important aspects of responsible AI (Dignum, 2018; Mikalef et al., 2022), encouraging “development of AI technology in a way that *secures people’s trust, serves the public interest, and strengthens shared social responsibility*” (Floridi et al., 2018, p. 701 (original emphasis)). Rather than designing responsible AI we suggest designing for *responsible use of AI*—such as chatbots—taking seriously that the technology itself can never be fully transparent or accountable and hence never act responsibly. It is the overall service that needs to be responsible. This implies that making human advisors available to answer questions that Anna cannot answer is a way of securing a responsible public service for the WA.

Our data shows that the chatbot occasionally makes mistakes and answers correctly to a wrong intent, i.e., a wrong interpretation of the question. The text in all the answers from the chatbot is formulated by experienced welfare advisors and are therefore always correct about the benefits. However, the matching of the requests with the readymade answers is done by ML, and this match is the topic of our analysis of the inner workings of Anna. Identifying if and how the ML machinery can produce weak or insufficient matching of questions and responses can point to where the ML can be improved as well as to where the limits for chatbot service could—and should—be set. Our adaptation of Suchman's framework for human-machine interaction to the chatbot conversations was useful for identifying examples where the citizen's questions and the chatbot's responses were not aligned. Supplementing this with Taylor's framework of information needs added a processual perspective on the chatbot conversations where the support for developing answerable questions is key.

8 Conclusion

In this paper, we have studied questions to Anna, the chatbot of the Norwegian welfare administration (WA) and analysed how the chatbot responds to actual questions from the citizens about welfare benefits. From logs of more than 8000 chat sessions, we have selected sessions where the citizen did not seem to get a good answer and analysed these to see how they met the citizens' information needs. A large number of chatbot conversations are successful, but we contend that analysing insufficient or misguided examples of chatbot conversations will contribute to understanding how the chatbot can be improved. Our selected examples document how and why mismatches between citizens' information needs and the chatbot's answers occur, which is particularly relevant in a discussion of fairness, transparency, and responsible AI in democratic governmental institutions.

We found that the chatbot gave incomplete or misleading information in response to the citizen's question in several chatbot sessions. Anna often responded with not understanding the question. All answers about the benefits are correct, but problems occur with the matching of information needs with answers. An erroneous match can lead to misunderstandings for the citizens about their eligibility for benefits. In some of the answers important information was omitted if the citizen did not ask for it. We also described an example where Anna did not detect an inconsistent question and ended up giving a correct answer to the wrong question. We noticed that many citizens wrote long questions as if talking to a human when chatting with Anna. However, long questions often contain several keywords, increasing the risk for an erroneous match

and therefore an irrelevant chatbot answer. We found that the *disguised level four questions* are particularly difficult for the citizen to assess as the answers seem relevant but do not include all relevant information. A good answer presupposes that the citizen asks good questions to the chatbot. By using buttons, the citizen can select among good answer alternatives presented by the chatbot, but we see this as a limited form of question negotiation supporting the citizen to arrive at a more precise question that will more likely receive a correct and relevant answer.

One contribution from this paper is to show how Suchman's framework helped us identify interaction mismatches where actions or information given by the citizen were lost. A second contribution is our description of information needs and how they do not come into a citizen's mind fully formed, but instead become increasingly well formulated during the conversation until they become questions that fit the chatbot's repertoire of keywords. Here we used Taylor's theory about the four levels of information needs to analyse the questions to and answers from the AI-based system, and the lost information was important for our interpretation of whether an answer was satisfactory. Our third contribution is to show how the inner workings of the chatbot can explain some of the chatbot's errors and omissions. We show how the inner workings of the chatbot shapes the answerable questions. In line with other critical research on AI-based systems we argue that identifying even one single mismatch which lead to a disguised level four response where we can document that the mismatch stems from the way the ML analysis of the question is enough to refute that the chatbot always answers with relevant, correct, and sufficient information (e.g., Antun et al., 2020; Broussard, 2018; Dignum, 2018; Mikalef et al., 2022).

Giving correct and relevant answers to citizens' information needs is a form of responsibility for a public institution. Responsible use of AI therefore needs to "balance between human- and machine-led decision-making ... [where] the AI should promote the autonomy of all human beings *and control* ... the autonomy of computer systems" (Floridi et al., 2018 p. 698) (original emphasis)). The AI-based chatbot answers many questions well. However, we argue that knowledge about AI's strengths and weaknesses—being realistic about its potentials as well as its limits — is a basis for designing a responsible overall public service that also includes AI. Knowledge about AI's limits implies including in the planning of the overall service that the chatbot may not answer all of its questions sufficiently. The fact that the WA make human advisors available to answer citizens who are not sure that the chatbot has given a sufficient answer is a way of addressing this limitation that maintains the WA's responsibility towards the citizens.

A main point of this paper is that a public service is accountable to the citizens and being a well-functioning digital public service depends on their experiences. This is

why we emphasize the problem of the *disguised level four questions*, where the citizen has no visible cue that something is missing. Providing some kind of transparency of the chatbot's inner workings available for the citizens may help them evaluate the chatbot's answers, e.g., as an option to see the intent representing the question Anna responded to. However, many citizens do not even read all the text in the answers and adding more text may not solve the problem. Finding ways of encouraging citizens to check which question was answered by the chatbot may be of help for some but may at the same time encourage mistrust in the chatbot's answers. Designing a responsible public service also means being realistic about what can be expected from the citizens using the service. People can have high expectations to AI and interpret Anna's personalized answers differently than they would interpret reading general information on a web page. In this paper we argue that helping the citizens to have realistic expectation of AI technology is crucial for designing a responsible public service in which AI has a role.

Notes

1. Since 2019, Anna has continuously been extended with more welfare domains and keywords.
2. For the chatbot trainers in WA, transparency is provided through the chatbot administrator panel.

Acknowledgment

We first want to thank our contacts in the public Welfare Administration for letting us in to investigate problems with their computer-supported services—we hope our studies can help them improve their services. Thanks to Gerd Berget, who suggested Taylor's theory to us. We also want to thank the reviewers and editors of SJIS for pushing us to improve the argumentation. Finally, thanks to Trenton Schulz for his help in improving the language.

References

- Aggeliki, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y., (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, 36(2), 358-367.

- Adam, M., Wessel, M. & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electron Markets*, 31, 427-445, <https://doi.org/10.1007/s12525-020-00414-7>
- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennet, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R. & Horvitz, E. (2019). *Guidelines for human-AI interaction*, In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-13).
- Antun, V., Renna, F., Poon, C., Adcock, B., & Hansen, A.C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, Dec 2020, 117(48), 30088-30095, DOI: 10.1073/pnas.1907377117
- Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, 22(2).
- Brandtzaeg, P. B., & Følstad, A., (2017). Why people use chatbots. *International conference on internet science*, Springer, Cham, pp. 377-392.
- Brandtzaeg, P. B, & Følstad, A., (2018). Chatbots: Changing User Needs and Motivations. *Interactions*, 25(5), 38-43.
- Bratteteig, T., & Verne, G. B. (2012). Conditions for Autonomy in the Information Society: Disentangling as a public service. *Scandinavian Journal of Information Systems*, 2012, 24(2), 51-78.
- Broussard, M. (2018). *Artificial unintelligence: how computers misunderstand the world*. Cambridge, MA, MIT Press.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, January-June: 1-12. <https://doi.org/10.1177/2053951715622512>
- Caldarini, G., Jaf, S. & McGarry, K. (2022). A Literature Survey of Recent Advances in Chatbots. *Information*, 13(41). <https://doi.org/10.3390/info13010041>

- Chang, Yu-Wei. (2013). The influence of Taylor's paper, Question-Negotiation and Information-Seeking in Libraries. *Information Processing & Management*, 49(5), 983-994.
- Chaves, A. P., & Gerosa, M. A. (2021). How Should My Chatbot Interact? A Survey on Social Characteristics in Human-Chatbot Interaction Design. *International Journal of Human-Computer Interaction*, 37(8), 729-758.
- Dibitonto, M., Leszczynska, K., Tazzi, F., & Medaglia, C. (2018). *Chatbot in a Campus Environment: Design of LiSA, a Virtual Assistant to Help Students in Their University Life*. Human-Computer Interaction, Interaction Technologies, HCI 2018, 103-116.
- Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology*, 20(1), 1-3. <https://doi.org/10.1007/s10676-018-9450-z>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28,689-707.
- Fryer, L., Ainley, M., Thompson, A., Gibson, A. & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, 75, 461-468.
- Grudin, J. & Jacques, R. (2019). *Chatbots, Humbots, and the Quest for Artificial General Intelligence*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Association for Computing Machinery, New York, NY, USA, Paper 209, 1-11.
- Holmquist, L. E. (2017). Intelligence on Tap: Artificial Intelligence as a New Design Material. *interactions*, 24(4), 28-33.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

- Jordan, M.I. & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kane, G. C., Young, A., Majchrzak, A. & Ransbotham, S. B. (2021). Avoiding an Oppressive Future of Machine Learning: A Design Theory for Emancipatory Assistants, *MIS Quarterly*, 45(1), 371-396.
- Kocielnik, R., Amershi, S., & Bennett, P.N. (2019). *Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Association for Computing Machinery, New York, NY, USA, Paper 411, 1-14.
- Liao, Q.V., Hussain, M.M., Chandar, P., Davis, M., Khazaen., Y., Crasso, M.P., Wang, D., Muller, M., Shami, N. S. & Geyer, W. (2018). *All Work and No Play? Conversations with a Question-and-Answer Chatbot in the Wild*. CHI 2018, pp: 3.1-3.13, Montreal, QC, Canada
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell Sage Foundation.
- Luger, E. & Sellen, A. (2016). *Like Having a Really Bad PA: The Gulf Between User Expectation and Experience of Conversational Agents*. CHI 2016, 5286-5297.
- Lyytinen K., Nickerson J.V., & King J.L. (2021). Metahuman systems = humans + machines that learn. *Journal of Information Technology*, 36(4), 427-445, doi:10.1177/0268396220915917
- Martin, D., O'Neill, J., Randall, D. & Rouncefield, M. (2007). How Can I Help You? Call Centres, Classification Work and Coordination. *Journal of CSCW*, 16 (3), 231-264.
- Mikalef, P., Conboy, K., Lundström, J. E. & Popovic, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. *European Journal of Information Systems*, 31(2), 257-268. DOI: 10.1080/0960085X.2022.2026621

- Muller, M. J. (1999). Invisible Work of Telephone Operators: An Ethnocritical Analysis, *Comput. Supported Coop. Work*, 8(1-2), 31-61.
- Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q.V., Dugan, C. & Erickson, T. (2019). *How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation*. CHI 2019, paper 126, 1-14.
- Parmiggiani, E., Østerlie, T. & Almklov, P. G. (2022). In the Backrooms of Data Science. *Journal of the Association for Information Systems*, 23(1), 139-164.
- Russell, S. & Norvig, P. (2010). *Artificial Intelligence: A modern approach*, Pearson, Boston.
- Savolainen, R. (2017). Information need as trigger and driver of information seeking: a conceptual analysis. *Aslib Journal of Information Management*, 69(1): 2-21, <https://doi.org/10.1108/AJIM-08-2016-0139>
- Simonsen, L. (2019) *Når brukerdialogen automatiseres— hva blir vanskelig?*, Master thesis, University of Oslo
- Simonsen, L., Steinstø, T., Verne, G. & Bratteteig, T. (2020). “I’m Disabled and Married to a Foreign Single Mother”. *Public Service Chatbot’s Advice on Citizens’ Complex Lives*. In: S. Hofmann et al. (Eds.), *Electronic Participation, ePart 2020, Lecture Notes in Computer Science, 12220*, 133-146, Springer, Cham. https://doi.org/10.1007/978-3-030-58141-1_11
- Shevat, A. (2017). *Designing bots: creating conversational experience*. O’Reilly: Beijing.
- Skaarup, S. (2021). *Beyond Substantive Goals — A Framework for Understanding Citizens Need and Goals in Bureaucratic Encounters*. In *Electronic Government: 20th IFIP WG 8.5 International Conference, EGOV 2021 Proceedings*. Springer-Verlag, Berlin, Heidelberg, 86-102. https://doi.org/10.1007/978-3-030-84789-0_7
- Steinstø, T. (2020). *Bruk av chatbot i praksis—En kvalitativ studie av utfordringer ved bruk av chatbot i offentlige tjenester*. Master thesis, University of Oslo.

- Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.
- Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. 2nd ed, Cambridge: Cambridge University Press.
- Taylor, R. (1968). Question-Negotiation and Information Seeking in Libraries. *College and Research Libraries*, 29(3), 178-194.
- Teodorescu, M. H. M., Morse, L., Awwad, Y. & Kane, G. C. (2021). Failures of Fairness in Automation Require a Deeper Understanding of Human-ML Augmentation. *MIS Quarterly*, 45(3), 1483-1500.
- Timmins, F. (2006). Exploring the concept of “information need”. *Int J Nurs Pract*, 12(6), 375-81. doi: 10.1111/j.1440-172X.2006.00597.x.
- Tyckoson, D. (2015). Question-Negotiation and Information Seeking in Libraries: A Timeless Topic in a Timeless Article. *College & Research Libraries*, 76(3), 247-250.
- Shollo, A., Aanestad, M., Lyytinen, K., Mahring, M., Nielsen, P. A. & Sørensen, C. (2020). *Panel 4: A Nordic Sociotechnical Perspective on Artificial Intelligence Phenomena*. Proceedings of the 28th European Conference on Information Systems (ECIS): A Virtual AIS Conference.
- von Wolff, R.M., Hobert, S. & Schumann, M. (2019). *How May I Help You? -- State of the Art and Open Research Questions for Chatbots at the Digital Workplace*, HICCS 2019, 95-104.
- Zamora, J. (2017). *I'm sorry, dave, i'm afraid i can't do that: Chatbot perception and expectations*. The 5th International Conference on Human Agent Interaction, 253-260.
- Verne, G. (2014). Two faces of autonomy. Learning from non-users of an e-service, *Systems, Signs and Actions*, Special issue on “Government—citizen communication through the web”, 8(1).

- Verne, G. (2015). *The winners are those who have used the old paper form. On citizens and automated public services*, PhD Thesis, Dept. of Informatics, University of Oslo.
- Verne, G., Oskarsen, J. & Bratteteig, T. (2022). *The human touch meets digitalization: on discretion in digitized services*, In: Electronic Participation, ePart 2022, Lecture Notes in Computer Science, vol 12222, Springer, Cham.
- Verne, G., Bratteteig, T. (2016). Do-it-yourself services and work-like chores: on civic duties and digital public services. *Pers Ubiquit Comput*, 20(4), 517-532. <https://doi.org/10.1007/s00779-016-0936-6>
- Walsham, G. (2002). Interpretive Case Study in IS Research: Nature and Method, in: Myers & Avison (eds.) *Qualitative Research in Information Systems*, Sage, 101-114.
- Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: Freeman.
- Wollny S., Schneider J., Di Mitri D., Weidlich J, Rittberger M. & Drachsler H. (2021). Are We There Yet?—A Systematic Literature Review on Chatbots in Education. *Front. Artif. Intell.*, 4,654924. doi: 10.3389/frai.2021.65

