



User Interactions With a Municipality Chatbot—Lessons Learnt From Dialogue Analysis

Asbjørn Følstad & Nina Bjerkreim-Hanssen

To cite this article: Asbjørn Følstad & Nina Bjerkreim-Hanssen (31 Jul 2023): User Interactions With a Municipality Chatbot—Lessons Learnt From Dialogue Analysis, International Journal of Human-Computer Interaction, DOI: [10.1080/10447318.2023.2238355](https://doi.org/10.1080/10447318.2023.2238355)

To link to this article: <https://doi.org/10.1080/10447318.2023.2238355>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 31 Jul 2023.



[Submit your article to this journal](#)



Article views: 1207



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

User Interactions With a Municipality Chatbot—Lessons Learnt From Dialogue Analysis

Asbjørn Følstad^a and Nina Bjerkreim-Hanssen^b

^aSINTEF, Oslo, Norway; ^bProkom, Oslo, Norway

ABSTRACT

Chatbots are increasingly taken up by the public sector, as a means to efficient provision of information and services. However, there is a lack of knowledge on how users interact with such chatbots. To address this knowledge gap, we have conducted an analysis of user interactions with a chatbot for citizens of Norwegian municipalities. We analyzed a total of 2663 user-chatbot dialogues from six municipalities, using the framework of Følstad and Taylor. The analysis showed that most user input was characterized by brief messages and a utility-oriented dialogue style whereas chatbot responses were characterized by substantial response relevance (68% of chatbot responses categorized as relevant) and helpfulness (66% of dialogues categorized as help being offered and likely used). Furthermore, message brevity and a utility-oriented dialogue style was found to be positively associated with users receiving relevant chatbot responses and helpful dialogue outcomes. Variation in chatbot design, specifically how the chatbot was presented to users, was found to impact user message brevity and dialogue style, and, by extension, response relevance and dialogue outcome. On the basis of the findings, we summarize lessons learnt and suggest directions for future research.

1. Introduction


The public sector is increasingly making use of chatbots in their interaction with citizens (Makasi et al., 2022). Chatbots are software agents that provide access to information and services through a conversational user interface (Følstad et al., 2021) and are seen as a potentially effective and efficient means for providing information and support (Makasi et al., 2022). In the public sector, chatbots have been taken up both for specific service areas, such as welfare services (Verne et al., 2022) and pandemic response (Amiri & Karahanna, 2022), and for access to a broader set of services, for entire city administrations (Van Noordt & Misuraca, 2019) or municipalities (Abbas et al., 2022). Chatbots in the public sector are typically provided for service triaging, that is, to help users navigate the public sector service offering and provide information on frequently asked questions (Makasi et al., 2022)

The current body of knowledge on chatbots in the public sector is limited. While initial studies have provided analytical evaluations (Van Noordt & Misuraca, 2019) and initial user insight through interview (e.g., Henk & Nilssen, 2021; Makasi et al., 2022) or questionnaire studies (Tisland et al., 2022), there is a scarcity of studies on how users interact with chatbots in the public sector. This is surprising, as

chatbots represent a fundamentally different approach to interaction with public sector information and services compared to what is provided through established self-service technologies or manned support centres. To understand the impact of chatbots in public sector service provision and pave the way for future advancement of such chatbots, it is important to establish knowledge about how such chatbots are used.

In the present study, we fill this knowledge gap by systematically exploring user interactions with a public sector chatbot. Specifically, we investigate how users approach interactions with such chatbots, how the chatbot responds to user requests in terms of relevance and helpfulness, and how dialogue outcomes may be impacted by variation in users' interactions and chatbot design. For this purpose, we have analyzed 2663 user dialogues with a chatbot supporting citizens of Norwegian municipalities. The studied chatbot is intended to help users access needed information and services online and covers a broad range of public sector service areas such as childcare and education, healthcare, city planning, and recreation. This makes the chatbot a valuable point of departure for an investigation of user's interactions with chatbots in the public sector. The analysis was conducted with basis in the framework of Følstad and Taylor (2021), where key analytical constructs concern the

CONTACT Asbjørn Følstad  asf@sintef.no  SINTEF, Oslo, Norway

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10447318.2023.2238355>

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

relevance and helpfulness of chatbot dialogues. The analyzed dialogues were sampled from a chatbot implemented in six different municipalities. Because these municipalities differed systematically in their presentation of the chatbot, the study also allowed for an exploration of the impact of such variation.

The study contributes significantly to the body of knowledge on chatbots in the public sector. The study provides novel insight into how users interact with a public sector chatbot, what can be expected from such chatbots with regards to response relevance and dialogue helpfulness, how users' message brevity and dialogue style may impact chatbot response relevance and dialogue outcome, and how the chatbot design may impact users' dialogue style. As such, the study has both theoretical and practical interest. Furthermore, on the basis of the study findings, we provide lessons learned to support future designs of chatbots for the public sector.

The remainder of this paper is structured as follows. First, we provide an overview of relevant background knowledge. We then present the research questions and detail the study method before presenting the study findings. In the subsequent discussion, we reflect on our findings relative to the research questions and related research, summarize implications and lessons learnt, and point out limitations and future research.

2. Background

Key background knowledge for this study includes details on public sector service provision, chatbots implemented for this purpose, user interaction and experience with chatbots, and dialogue analysis as a means for studying chatbot dialogue. We will provide this in the following subsections. The background overview has been established as part of a 3-year international research project.

2.1. Digital public sector services

Public sector service provision concerns rights-based information and services from government agencies (Stahl, 2005) and covers a range of services within areas such as education, healthcare, infrastructure, welfare, and regulation. Because government services are based on citizens' democratic rights, it is particularly important that these are provided in a manner that satisfies legal obligations and user expectations. Public sector services are provided through a complex organization at local and state levels, and across service areas, by specialized interlinked organizations. Because of the broad span in public sector services and the complexity in the underlying service organizations, it may be challenging for citizens to understand where and how to get needed information and services.

Public sector service provision has strongly benefitted from digitalization. Government information and services to citizens are increasingly provided online by default and self-service is the preferred mode of service provision. The turn towards digital self-service is grounded in strategic policies,

e.g., in Europe (EC, 2022) and the US (US Department of State, 2012), to strengthen service access, efficiency, and quality.

Increasingly, digital public sector services are considered a means to strengthen the public value of government services, that is, a means to better satisfy the users of the services (Panagiotopoulos et al., 2019). Twizeyimana and Andersson (2019), in their literature review of public value of digital public sector services, found such value to concern improved public services, improved administration, and improved social value. For AI services (Fatima et al., 2022) and chatbots (Makasi et al., 2020) specific public values mentioned in the literature include service improvement, effectiveness, efficiency, transparency, trust, and accountability.

Concern for public value in digital public sector services have, predictably, been found to impact citizen uptake. For example, Carter and Bélanger (2005), in a questionnaire study, found compatibility, ease of use, and perceived trustworthiness to predict intention to use digital public sector services. Furthermore, the awareness of the public value of digital public sector services have motivated increased attention to their ethical and societal implications, with an increased attention towards the trustworthiness of digital technology (European Commission [EC], 2022), and its ethical development (Hansson & Fröding, 2022).

While digitalization has motivated an external orientation of public sector service provision, with benefits for public value, digitalization has also increased the complexity and specialization of such services (Janowski, 2015). Hence, an enduring challenge in digital public sector service provision concerns the efficiency and effectiveness of service provision, including support for users to identify where to find needed information and service and which public sector organization that provides this.

2.2. Chatbots in public sector service provision

Chatbots have been taken up to support interaction between public sector service providers and citizens at all levels of service provision. Chatbots may be particularly helpful as a point of contact for public sector service provision, given this sector's broad range of services and information sources distributed across a complex organization (Androutopoulou et al., 2019). A report published by the EC (2020), identified 52 examples of public sector chatbots across Europe. These were seen as having the potential to bring about incremental changes, e.g., answering commonly asked questions, or more disruptive or transformative changes, e.g., providing personalized service or helping citizens locate information from multiple government agencies. The COVID-19 pandemic has driven increased interest in chatbots as a means for public sector service provision, due to the need for efficient means of providing information and responding to citizen requests on COVID-19-related policies and regulations, health advice, vaccination, and testing (e.g., Amiri & Karahanna, 2022).

Chatbots in the public sector are typically set up as retrieval-based solutions where machine learning is used to predict user intents, based on the users' free text input, and

return the predefined action corresponding to this intent. Intent prediction is constituted by what McTear (2021) refers to as statistical data-driven systems based on machine learning, whereas requests are fulfilled through rule-based provisioning and refinement of chatbot responses.

Makasi et al. (2022) noted that public sector service providers may take up chatbots to achieve internal objectives, e.g. reduce service delivery cost and improve efficiency, and external objectives, e.g. strengthen service access and improve service experience. However, for chatbots to contribute towards these objectives, their design and sophistication needs to match the specific public sector service use case. Towards this, Makasi et al. proposed a typology of public sector chatbots, distinguishing between chatbots for *service triaging*, helping users access needed information and services, *service information gathering and analysis*, helping users with adapted information and services, and *service negotiation*, helping users with personalized information and service provision. Makasi et al. noted that while there are relatively few examples of chatbots at the levels of service negotiation or service information gathering, chatbots at the level of service triaging are more widespread. This may in part be due to the relative immaturity of chatbots in the public sector, and in part due to the requirements for system integration required for more sophisticated chatbots.

Chatbots for service triaging are typically implemented to provide generic information without access to user profiles and may be beneficial to handle frequently asked questions. Vassilakopoulou et al. (2023), in a case study of chatbot implementation in the Norwegian welfare administration, found that chatbots may be beneficial for answering of generic routine questions, access to updated information, and provision of monitoring capabilities to identify service disruptions. Likewise, Henk and Nilssen (2021), in a case study on experiences from implementation of a chatbot as part of the customer centre of a Scandinavian public sector service provider, reported positive experiences on the chatbot's ability to handle routine questions while noting that human service providers were needed to handle complex or personal requests.

Van Noordt and Misuraca (2019), in a review of three public sector chatbots in different European countries, noted a relatively low level of complexity in the provided services and concluded that these are typically implemented to strengthen information provision, with limited provision of transactional services and integrated information. They also noted the need for empirical research to better understand chatbot implementations within the public sector.

2.3. User experience of chatbots in the public sector

As called for by Van Noordt and Misuraca (2019), research efforts to understand uses and opportunities associated with chatbots in the public sector have increased. Specifically, there has been an increase in research on how users experience such chatbots. This research has typically been conducted as interview or questionnaire studies, to gain insight into user's self-reported perceptions of chatbots in public

sector service provision. Abbas et al. (2022), in an interview study on users' intentions to take up a municipality chatbot, provided insight on how performance expectations, effort expectations, and trust are key for such chatbot uptake. Makasi et al. (2022), in an interview study with public sector chatbot users and designers, identified a set of interlinked values associated with such chatbots, including effectiveness, efficiency, accessibility, ease of use, and predictability. Tisland et al. (2022), in a questionnaire study involving e-government chatbot users, found information quality (e.g., meeting users' needs and being accurate) and service quality (e.g., being accessible and dependable) to be particularly important determinants of users' trust in such chatbots and, consequently, their satisfaction and intention to use. Finally, Aoki (2020), in a questionnaire study on chatbot service concepts in the public sector, found that users' tendency to trust public sector chatbots depends on the public sector domain characteristics as well as the perceived reason for introducing the chatbot.

There has, however, been limited attempts at investigating user experience through analysis of how users interact with public sector chatbots. An exception to this lack of research is a series of studies conducted on the chatbot of the Norwegian welfare administration. Here, chatbot dialogues were analyzed to identify causes of chatbot response failure. Simonsen et al. (2020) found such failure to be caused by the chatbot not recognizing key concepts or inconsistencies in the user request, or not being able to map the user request to the correct service area. Verne et al. (2022), in a continuation of the study by Simonsen et al., found that a key reason why the studied chatbot failed was an inability to help users advance from imprecise requests to requests fitting the conceptual framework of the service provider. To our knowledge, no studies have been presented based on dialogue analysis of general user interactions with public sector chatbots.

2.4. Dialogue analysis to investigate chatbots

Users' interactions with chatbots leave detailed logs which provide potentially rich insight into user goals and the chatbot's ability to address these. On this background, the lack of systematic studies based on chatbot log data is surprising. In particular, as the majority of chatbots in public sector service provision are at the level of information triage (Makasi et al., 2022) with no user profile generation or person data gathered.

For the analysis presented in this paper, we will draw on Følstad and Taylor (2021) framework for chatbot dialogue analysis, drawing on linguistic theory (Sacks et al., 1974; Schegloff, 2000) and its adaptation to understand and design for conversational user experience (Moore, 2018). The framework was developed to investigate chatbots for customer service, but it is arguably well suited also for the public sector because of the common concern for effectiveness and efficiency in chatbot interaction within public sector service provision and customer service. Furthermore, the framework addresses interactions with intent-based

customer service chatbots, based on the same technology as that of current public sector chatbots. The framework has been developed across several studies in collaboration with chatbot hosts and has been used for applied and theoretical purposes.

In the framework, dialogue analysis is conducted on the log from chatbot dialogues. That is, the trails of the interaction sequences between a user and a chatbot. This log typically contains details on time, number of messages, user free text input, predicted user intents, and associated chatbot responses. The user free text inputs and chatbot responses have the form or messages in the dialogue. Predicted user intents concerns the mapping of the user's free text input to the chatbot's predefined set of possible users' intentions.

The textual exchanges between the users and the chatbot provide insight into the user goals, as reported in free text by the users, the corresponding chatbot responses, and the subsequent refinements of the dialogue through users' selections of predefined follow-up alternatives or follow-up free-text messages.

The dialogue analysis is conducted as a manual process. This makes it possible to identify false positive and negative responses from the chatbot, that is, responses that do not provide good matches between the actual intents of the users' messages and the responses offered by the chatbot. With basis in linguistic theory (Sacks et al., 1974; Schegloff, 2000), the units of analysis are message sequences and entire dialogues. A message sequence, constituted by an adjacency pair of a user message and a corresponding chatbot response, is analyzed for response relevance and understandability (Moore, 2018). A dialogue, which entails all the messages in the chat session and may be constituted by one or more message sequences, is analyzed for dialogue outcome and efficiency.

Other frameworks have also been provided for the analysis of chatbot dialogue, including analysis of the specificity and relevance of open-domain chatbot answers (Adiwardana et al., 2021) and the much-cited PARADISE framework for evaluation of spoken dialogue conversational agents (Walker et al., 1997). However, the framework of Følstad and Taylor, developed for dialogue analysis in the customer service domain, arguably is a better fit for analysis of interactions with public sector chatbots as such interactions closely resembles those of chatbots within customer service.

3. Research questions

Motivated by the gap in current knowledge on how users interact with chatbots in the public sector, we set up the following general research question:

How do users interact with a public sector chatbot?

Specifically, we were interested in four aspects of this question. First, we wanted to explore users' input to public sector chatbots, that is, the content of the users' messages and their dialogue style. Second, we wanted to analyze the chatbot output in response to the users' messages, that is, whether and how these supported the users' stated goals of the interaction. Third, we wanted to study the impact of the

user messages and dialogue style on the chatbot output, that is, whether and how users' formulation of their input impacted chatbot responses. Fourth, we wanted to investigate the impact of the chatbot design on the user input, that is, whether and how changes in the chatbot may impact how users interact with the chatbot. In consequence, the research question could be broken down in four sub-questions:

- RQ1. How do users approach interactions with public sector chatbots?
- RQ2. How does a public sector chatbot respond to users' requests in terms of relevance and helpfulness?
- RQ3. How does the users' approach to the chatbot interaction impact response relevance and helpfulness?
- RQ4. How does the design of the chatbot impact users' approach to chatbot interaction.

The general research question, hence, guided research explorations of both the users' (RQ1) and chatbot's (RQ2) contributions to the interaction, and how these contributions might impact each other (RQ3 and RQ4). As such, the explorations could provide insight of theoretical importance, on users' interaction patterns when using a public sector chatbot, and of practical importance, on how chatbot design choices may impact users' interactions and, ultimately, chatbot response relevance and helpfulness.

4. Method

In the method section we will first provide an overview of the research design before detailing our choice of chatbot for the study, our approach to dialogue sampling and analysis, as well as research ethics.

4.1. Research design

In response to the research question, we conducted an analysis of user interactions with a municipality chatbot. User interactions were recorded in the logs of the chatbot, and data were sampled to provide a balanced exploration of user interactions across variations in available municipalities. The analysis was based on the framework of Følstad and Taylor (2021). In the following, we detail the choice of chatbot for the study, our approach to dialogue sampling and analysis, as well as the ethical considerations for this study.

4.2. Choice of chatbot

The studied user interactions were conducted with the municipality chatbot called *Kommune-Kari*. The chatbot is available to citizens in more than 100 municipalities in Norway and has also been taken up by municipalities in other Nordic countries. The chatbot is based on the boost.ai conversational platform and provides access to information and services from the municipality through text-based dialogue. It is a chatbot for service triage, as understood in the classification of Makasi et al. (2022), and it provides generic

responses on user requests without access to user information or profiles. The chatbot may be accessed through the respective municipalities' websites. The interaction starts out with an introduction message from the chatbot where it presents itself as the municipality chatbot and encourages the user to enter their request in a brief manner and not to provide any personal information. Users enter their requests as free text messages with a max length of 110 characters.

The *chatbot design* differs between municipality websites in the sense that its visual presentation is varied. In some websites, it is presented centre-stage of the landing page, something that makes it highly visible to users, in addition to being available from other pages on the website as well. In other websites, it is presented as an avatar in the lower right-hand corner of the landing page as well as other pages of the website. When the chatbot is activated by the user, a dedicated on-screen element for the dialogue interaction is opened on the right-hand side of the screen—regardless of how it was presented to the user. Apart from this systematic variation in visual presentation, the chatbots were highly similar in design and structure. For example, in terms of the size and layout of the chat dialogue, in terms of chatbot conversation style, and in terms of the presented content. Variation in chatbot design is illustrated in Figure 1.

The chatbot covers a broad range of municipality information and services, on topics such as health, education, infrastructure, leisure, planning and regulation. If an answer to a request is simple and concrete, it may be provided directly in the chatbot response. Otherwise, the answer may be provided by reference to further information or services on the municipality website. Such references are made through links or buttons in the chatbot's responses.

The chatbot is intent-based, meaning that it accepts free-text messages from users. From these messages, the users' intents are predicted based on a machine learning model. Due to the broad scope of Norwegian municipality information and services, the chatbot prediction model includes about 6000 intents. When a user intent is identified, a corresponding action is provided in the form of a chatbot response. The chatbot responses are predefined and provided in a rule-based fashion. As part of its response, the chatbot may offer options for refinement of the request through interactive buttons or follow-up information.

The chatbot also includes mechanisms for conversational repair, that is, response to troubles preventing a

conversation from progressing (Moore, 2018). In cases where no intents reach sufficient prediction confidence, the chatbot initiates conversational repair by returning a fallback answer asking the user to rephrase or inviting the user to contact manned support. In cases where several candidate intents reach above prediction confidence and have relatively similar confidence scores, the chatbot returns an uncertainty response, also to initiate conversational repair, where it explains to the user that it is uncertain of what the user wants and provides candidate paths for further interaction which reflect the most likely predicted intents.

Because different municipalities have overlapping information and service obligations, the chatbot Kommune-Kari is hosted by a common service provider, Prokom. This provider is responsible for maintenance and updates to the machine learning model and the chatbot responses. The chatbot has been in use since 2017. At the time of this study, it was engaged in about one million dialogues each year. Because of the relatively long period of deployment and substantial use, Kommune-Kari is a useful context for investigating user interactions with chatbots in the public sector.

4.3. Dialogue sampling

Dialogue data was sampled strategically to enable a balanced exploration of users' chatbot interactions. Specifically, we sampled dialogues from municipalities of varying size and geographical location. Furthermore, to allow for systematic explorations of the impact of chatbot design, we sampled dialogues both from municipalities where the chatbot was presented centre-stage of the municipality website landing page and from municipalities where the chatbot was presented as an avatar in the lower right-hand corner of each webpage.

As the dialogue analysis requires familiarity with the information and services provided in each given municipality, it was not feasible to include data from the more than 100 municipalities using the chatbot. Hence, dialogues were sampled from the chatbots of six different municipalities; three with the chatbot presented at the centre-stage on the municipality landing webpage and three with the chatbot presented in the lower right-hand corner of each webpage. The municipalities were strategically sampled to reflect variation in size and geographical location, while at the same

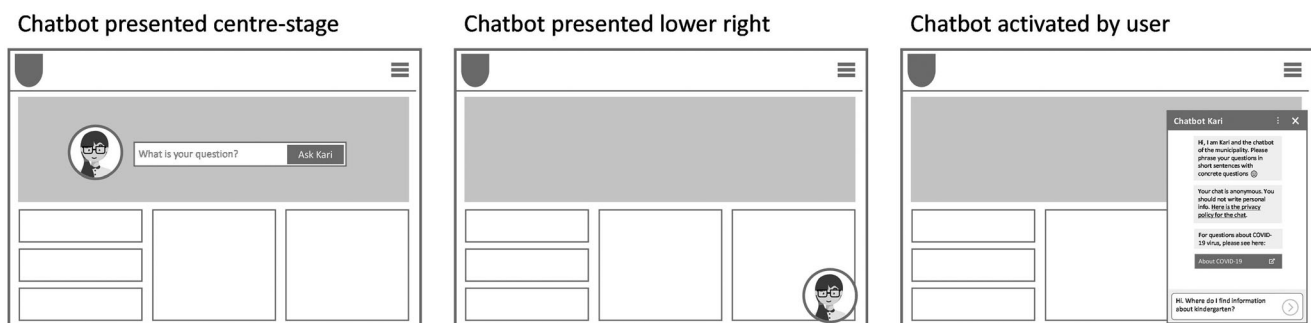


Figure 1. Variation in chatbot design. Chatbot presented centre-stage (left) or lower right (middle) of the municipality landing page. For both designs, the chatbot when activated by the user is contained in a dedicated chat element on the right hand of the webpage (right).

time ensuring that all participating municipalities had chatbots with sufficient volumes of use to enable sampling within the same time frame.

For sufficient breadth, we sampled between 400 and 500 dialogues from each municipality—in total 2663 dialogues. To be included in the sample, each dialog should include at least one meaningful free text user message. The log file for each dialogue included information on dialogue time and date, all user messages, predicted user intents, and corresponding chatbot responses, as well as user clicks on buttons or links provided by the chatbot. The dialogues were sampled June–July 2021. For chatbots with 400–500 dialogues in the sampling period, we sampled all dialogues. For chatbots with more than 400–500 dialogues in the sampling period, we sampled a subset of dialogues defined by a shorter time frame.

4.4. Analysis

The analysis was conducted with basis in the framework of Følstad and Taylor (2021) and consisted of an initial descriptive overview of the dialogues in terms of user requests and follow-ups, predicted intents, and chatbot responses. Following this, the dialogues were analyzed on the level of message sequences and entire dialogues.

At the level of *message sequences*, the material was analyzed with regards to response relevance and understandability. *Response relevance* concerns whether the chatbot response is a relevant response to the user message and entails the following codes: *relevant response* (the response is relevant for the user message), *false positive* (the response is irrelevant for the user message), *false negative* (the response is erroneously indicated to be out of scope), *out of scope* (the response is correctly indicated to be out of scope), *uncertainty response*, *relevant alternative* (the response indicates uncertainty in interpretation and offers alternatives, one of which is relevant), and *uncertainty response, no relevant alternatives* (the response indicates uncertainty in interpretation and offers alternatives, none of which are relevant). *Understandability* concerns whether the message sequence indicates any understandability problems for the user and involves the codes *likely understandable* (no indication of understandability issues) and *understandability issue* (indication in the dialogue of the user failing to understand the chatbot response).

Because the analysis of response relevance is of particular interest to our study, we provide examples of the different coding categories from the dialogues data in the study in Table 1.

At the level of *entire dialogues*, the material was analyzed with regards to dialogue outcome and dialogue efficiency. *Dialogue outcome* concerns whether the user's stated aim for the interaction was achieved and involves the following codes: *Relevant help likely used* (the support provided by the chatbot corresponds to the users stated aim and is provided either directly in the chat or through interactive elements such as buttons or links that has been used by the user), *relevant help likely not used* (the chatbot has provided access

to support through interactive elements, but these have not been used by the user), *escalation offered* (the chatbot offers escalation to manned customer service due to failure to provide needed support or due to municipality policy), *no relevant help* (the chatbot has not offered help of relevance to the user's stated aim). *Dialogue efficiency* concerns the progress in the dialogue and involves the following codes: *Coherent dialogue flow* (the dialogue is completed without breaks due to thematic change or comprehension failure), *breaks in dialogue change* (the dialogue includes one or more thematic changes or breaks due to comprehension failure).

Users' input to the dialogues was analyzed in terms of message brevity and dialogue style. *Message brevity* was analyzed by counting the characters in the first user message that stated the users' aim—typically the initial user message. *Dialogue style* concerns the use of greetings and displays of politeness (e.g., “hi” and “thank you”) and first and second pronouns in user messages. Dialogues containing both such identifiers were classified as reflecting a *social dialogue style* and dialogues not containing both such identifiers were classified as reflecting a *utilitarian dialogue style*.

Impact of chatbot design was analyzed by taking advantage of the fact that three of the municipalities from which dialogue data were sampled presented the chatbot centre-stage on the municipality landing page whereas three presented the chatbot in the lower right-hand corner of the website. The two modes of chatbot design were made subject of direct comparisons at the level of message sequences and entire dialogues as well as users' message brevity and dialogue style. The dialogue analysis was conducted by the first author and two chatbot AI trainers at Prokom, the provider of Kommune-Kari, one of the AI trainers being the second author. The two chatbot AI trainers analyzed 17% of the dialogues together, involving the first author to clarify any uncertainties in coding. On this basis, the remaining dialogues were analyzed by the first author.

4.5. Ethics

Ethical considerations were important to the planning of the study, as it concerns analyzing users' actual interactions with a public sector chatbot. Ethical aspects were assessed in a structured manner through the ethics appraisal process of the research project of which the study was part. Furthermore, given the importance of data protection for the research, the research design was discussed and followed up with the data protection officer of the involved research team. Also, the sampling of chatbot log data for research was conducted in line with descriptions for such sampling provided in the chatbot terms of use.

The sampling was conducted in collaboration with the municipalities as data controller and the chatbot host, Prokom, as data processor. Because users' free text requests to a chatbot, in principle, may contain personal information, despite the chatbot's request for users to not include such information, precautionary mechanisms were in place to ensure user anonymity. First, the chatbot platform

Table 1. Examples of coding categories for response relevance.

Codes—response relevance	Example message sequence
Relevant response	<i>User:</i> Want to book covid vaccine <i>Chatbot:</i> Here you can find information about vaccination in the municipality. [...] Button: Vaccine information. Button: Registration / invitation.
False positive	<i>User:</i> Possible to have vaccination in another municipality? <i>Chatbot:</i> You are considered protected if you are fully vaccinated [...]
False negative	<i>User:</i> Who can I contact regarding power infrastructure in my street. <i>Chatbot:</i> I do not understand. Maybe you want to ask in a different way? [...]
Out of scope	<i>User:</i> Is there miniature golf at [location] <i>Chatbot:</i> I do not understand. Maybe you want to ask in a different way? [...]
Uncertainty response, relevant alternative	<i>User:</i> Phone number to book vaccination <i>Chatbot:</i> Here are some suggestions which may be helpful: Button: Phone. Button: Covid vaccination.
Uncertainty response, no relevant alternatives	<i>User:</i> I am a student and want to take the second [vaccination] dose at my place of study. <i>Chatbot:</i> Maybe this can be of help? Button: Student

All example message sequences translated from Norwegian by the first author.

automatically masked any part of messages predicted to be a name, phone number, email address, or social security number. Second, personnel at the chatbot host manually sifted through all dialogues considered for sampling to filter out any dialogues containing person information. Hence, all analyses were conducted on anonymous log data. Finally, the findings of the study were fed back to the chatbot host and participating municipalities for insight to drive future service improvements.

5. Results

In the results section, we first detail findings concerning the characteristics of the user input to the chatbot (5.1) and the chatbot responses (5.2). Following this, we provide an overview of findings on how user input characteristics may impact chatbot responses (5.3) and, finally, how chatbot design may impact user input characteristics (5.4).

5.1. Characteristics of user input

The analyzed user input was made as part of 2663 dialogues, including 3610 message sequences. The vast majority of dialogues (77%) included only a single message sequence. That is, most users provided only one free text message in the dialogue to state their request. Some dialogues did, however, include more than one user message; 15% of the dialogues included two message sequences and 8% included three or more. See [Table 2](#) for examples of dialogues with single and multiple message sequences.

User messages tended to be formulated in a brief manner. Median length for the messages that contained the users' initial requests was 19 characters (25th percentile = 12; 75th percentile = 36) and 92% of these messages were 60 characters or less. Typical examples of such user messages were: "borrow walking aid," "technical department," "unnecessary noise from neighbor," and "where is health station youth".

The user input reflected an overwhelming utilitarian dialogue style; that is, it included relatively little greetings and displays of politeness and/or uses of first or second person pronouns. In total, 95% of the dialogues were coded as reflecting a utilitarian style. Only 5% were coded as

reflecting a social style. See [Table 3](#) for examples of dialogues in a utilitarian and social style.

As might be expected from the tendency towards message brevity and a utilitarian dialogue style, the users' requests to the chatbot typically concerned the main municipality service offerings. The most frequent areas of enquiry were (a) health-care, including requests related to the COVID-19 pandemic (30%), (b) municipality and contact information, including specifics on the municipality administration and contact to departments or persons (12%), (c) leisure, including request concerning arrangements and places of interest (6%), (d) case processing, including requests concerning applications and processing details (4%), (e) water and sewage, including error reports and status requests on water and sewage infrastructure (4%) (f) education, including requests concerning schools and school administration (3%), and (g) renovation, including requests concerning renovation pick-up and deposit points (3%). For examples of dialogues concerning the different areas of enquiry, see [Appendix 1](#).

We also analyzed the requests for user input indicating exploratory chatbot use, that is, users asking for information and support outside the scope of public sector service provision or users providing messages reflecting harassment or play. Such exploratory use was nearly absent in the material. Only 1.5% of the analyzed dialogues included exploratory user input. Some users, however, made use of the chatbot for questions which are relevant for public sector service provision, but which may be outside the scope of the municipality. Examples of this included requests concerning testing for COVID-19 prior to international travel (a service provided by private health companies, not the municipality) or requests concerning drivers licenses (a service provided by a state department, not the municipality).

5.2. Characteristics of chatbot responses

The chatbot responses were analyzed at the level of message sequences and at the level of entire dialogues. At message sequence level, the chatbot responses were analyzed for response relevance and understandability. At dialogue level, the chatbot responses were analyzed for dialogue outcome and efficiency.

Table 2. Examples of chatbot dialogues consisting of single and multiple message sequences.

Example dialogue—single message sequence	Example dialogue—two message sequences
<p><i>User:</i> When does one get a response on a corona test? <i>Chatbot:</i> Usually, 1-4 days are required to get the corona test report, but local variations occur. You will be notified on SMS or email when your test report is ready. If the test is positive (that is, you are inflicted by the corona virus) you need to take on home isolation.</p>	<p><i>User:</i> Opening hours <i>Chatbot:</i> Which opening hours do you need? Button: Citizen service. Button: Welfare service. Button: Recycling station. Button: Swimming pools. [...] <i>User:</i> Licensed premises. <i>Chatbot:</i> Here is information about licensed premises: Button: Registry of licenses. Button: License application. Button: Hours for licensed premises. <i>User (clicks button):</i> Hours for licensed premises. <i>Chatbot:</i> Licensed premises can be open but not allow new visitors after 24:00. [...]</p>

All example message sequences translated from Norwegian by the first author.

Table 3. Examples of chatbot interactions reflecting a utilitarian vs. social dialogue style.

Example dialogue—utilitarian style	Example dialogue—social style
<p><i>User:</i> Status for vaccination in [municipality] <i>Chatbot:</i> We provide the latest updates on vaccination status here: [link]</p>	<p><i>User:</i> Hi. Where in the municipality may I take a rapid test? <i>Chatbot:</i> In the first link below, you find information about corona test centres in [municipality]. Button: Corona test centre [...]</p>

All example message sequences translated from Norwegian by the first author.

5.2.1. Analysis at message level

At message level, the majority of chatbot responses of the entire dialogues were found to be relevant (59%) and less than a sixth were false positives (13%). Nearly a quarter of the responses were uncertainty responses (23%), some of these including a relevant answer option (9%) and some not (12%). For examples of coding at message level, see [Appendix 1](#).

Considering only the message sequences containing the users' initial formulations of their request, the proportion of relevant responses were somewhat higher (65%), with an additional set of uncertainty responses with a relevant answer alternative (10%). Hence, 75% of the chatbot's responses to the users' initial formulation of their request were relevant or contained a relevant alternative. Details are provided in [Table 4](#).

The chatbot responses were also analyzed for understandability, that is, the absence of indications of understandability issues. Across the analyzed data, 97% of the dialogues did not include message sequences suggesting understandability issues. Among the 3% of dialogues found to contain understandability issues, this typically concerned (a) the user not taking advantage of alternatives provided as buttons, suggesting this option to be overlooked, or (b) the user providing requests in a form not understandable to the chatbot, suggesting an overestimation of the chatbot's conversational intelligence.

5.2.2. Analysis at the level of entire dialogues

The dialogue outcome was mostly found to be helpful in guiding users towards their stated goals. In nearly two thirds of the dialogues, help was provided and likely used (65%), and in an additional set of dialogues, help was provided but not used (9%). In the remainder of the dialogues, users were provided either no help (21%) or directed towards manned service options (5%). Details are provided in [Table 5](#).

The chatbot dialogues were also analyzed for efficiency, concerning whether the dialogue contained breaks due to misinterpretations or failure to advance towards the user's stated goal. The vast majority of the dialogues (89%) did not

include such breaks. The remaining 11% of the dialogues included one or more breaks. Typically dialogues with such breaks were associated with issues in response relevance or dialogue outcome. For examples of dialogues reflecting the different dialogue outcomes, see [Appendix 1](#).

5.3. Impact of user input characteristics

While user input tended to be brief and utility-oriented, some users deviated from this pattern. We explored the implications of such variation in user input characteristics.

5.3.1. Impact of user message brevity

First, we analyzed the impact of users' brevity in phrasing their requests. For this analysis, we only included the messages containing the users' initial formulation of their requests. To provide a simple overview, we grouped these messages according to the following message length brackets: 1–20 characters, 21–40 characters, 41–60 characters, 61–80 characters, and 81–110 characters. We then analyzed response relevance for each of these brackets. Examples of user messages reflecting the different message length brackets are provided in [Table 6](#).

We found that the briefer user messages were more likely to receive a relevant response from the chatbot, and less likely to receive an uncertainty response, than were the longer user messages. The significance of this tendency was investigated in Chi-square tests for pairwise comparison of adjacent message length brackets. In the pairwise comparisons of relevant response (yes/no), Chi-square values indicated significant differences ($p < .01$) between all adjacent message length brackets except one. In the pairwise comparisons of uncertainty response (yes/no), the pairwise difference between message length brackets was significant ($p < .01$) between the first and second length bracket for comparisons concerning uncertainty responses with a relevant alternative, and between the first and second length bracket as well as the second and third for comparisons concerning

Table 4. Analysis of response relevance, detailed findings for frequency overall ($n = 3610$) frequency initial message ($n = 2663$).

Response relevance	Description	Freq. overall	Freq. initial
Relevant response (RR)	The response is relevant for the user message	59% (2441)	65% (1721)
False positive (FP)	The response is not relevant for the user message	13% (477)	11% (285)
False negative (FN)	The response erroneously indicates the user message to be out of scope	2% (88)	2% (50)
Out of scope (OoS)	The response correctly indicates the user message to be out of scope	4% (135)	2% (66)
Uncertainty response, relevant alternative (U-R)	The chatbot expresses uncertainty and offers alternatives, one relevant for the user message	9% (339)	10% (267)
Uncertainty response, no relevant alternatives (U-NR)	The chatbot express uncertainty and offers alternatives, none of which are relevant for the user message	12% (430)	10% (274)

Table 5. Analysis of dialogue outcome, detailed findings of frequency overall ($n = 2663$).

Dialogue outcome	Description	Freq.
Relevant help likely used	The chatbot offered support corresponding to the assumed user goal, either directly in the chat or through button / link alternatives used by the user.	65% (1731)
Relevant help likely not used	The chatbot offered support corresponding to the assumed user goal through button / link alternatives that were not used by the user.	9% (242)
Escalation offered	The chatbot offered escalation to other means of service, e.g., manned customer service due to interpretation failure or to policy.	5% (141)
No relevant help	The chatbot offered insufficient help towards the assumed user goal. No escalation was offered.	21% (549)

uncertainty responses with no relevant answer alternatives. Details are provided in Table 6.

It should be noted that while briefer messages tended to receive more relevant responses, the responses provided to the briefer messages could have a more general character. This may be illustrated by considering the user message examples for the shortest and longest text length brackets in Table 6, that is, “Change appointment” (shortest) and “Hi. I have an appointment for vaccination on Friday, but need to change the time. I do not find where I can do this” (longest). Both messages expressed the need to change an appointment. However, the briefest message, “change appointment,” received a higher-level response where the user first needed to clarify the kind of appointment the user wanted to change by selecting one of a predefined set of alternatives. In contrast, the less brief (and more precise) message, received a lower-level response where the chatbot already had established that the appointment concerned vaccination, and invited the user to specify which kind of vaccination the appointment change concerned.

5.3.2. Impact of user dialogue style

The impact of user dialogue style was investigated by comparing the dialogue outcome for interactions with a utilitarian style to the dialogue outcome for interactions with a social style. While only a small proportion of the dialogues reflected a social dialogue style (5%), the analysis nevertheless is theoretically interesting.

We conducted this analysis by considering the dialogue outcome *Relevant help likely used*. We chose to focus on this specific dialogue outcome as it arguably is the outcome most desirable to the users. Across the data from the six

municipalities, we found that the users engaging with the chatbot in a utilitarian style were more likely to receive relevant help and likely use it (1667 of 2528; 66%) than users engaging with the chatbot in a social style (64 of 135; 47%). The difference between the groups was found to be significant in a Chi-square test ($X^2 = 19.4, p < .01$).

The reason why a social dialogue style tends to receive a less desirable outcome may have several causes. In part, such a dialogue style may indicate more personal requests which may be challenging to support in a chatbot for anonymous use. In part, such a dialogue style may be associated with dialogues reflecting unrealistic expectations in the user in terms of chatbot conversational intelligence.

5.4. Impact of chatbot design on user input characteristics

The impact of chatbot design on user input characteristics was investigated by comparing user input for chatbots presented centre-stage of the municipality website with that of chatbots presented in the lower right-hand corner of the municipality website. Comparisons were conducted for user message brevity and user dialogue style.

5.4.1. Impact of chatbot design on user message brevity

The impact of chatbot design on user message brevity was investigated by comparing message length for the users' first messages containing the initial formulation of their requests. Users were found to formulate longer messages when the chatbot was presented in the lower right-hand corner of the website (median = 28; 25th percentile = 16; 75th percentile = 45) than when the chatbot was presented centre-stage

Table 6. Frequency of chatbot responses coded as relevant response (RR), false positive, (FP), false negative (FN), out of scope (OoS), uncertainty response, relevant alternative (U-R), and uncertainty response, no relevant alternatives (U-NR), for the different user message length brackets ($n = 2663$). Asterisks indicate a significant pairwise difference ($p < .01$) between the marked score and the score in the cell immediately below.

Message length	Count	Example user message	RR	FP	FN	OoS	U-R	U-NR
1–20 characters	1396	Change appointment	80%* (1123)	8% (116)	1% (19)	3% (38)	4%* (61)	3%* (39)
21–40 characters	748	Larger container for non-recyclable waste	58%* (437)	12% (88)	1% (9)	2% (12)	14% (102)	13%* (100)
41–60 characters	297	Where do I apply for renewed license to serve liquor	38% (112)	17% (50)	3% (9)	1% (4)	17% (51)	24% (71)
61–80 characters	118	Drive a car from Norway to Lithuania, what kind of covid test do I need?	30%* (35)	14% (16)	3% (4)	3% (3)	20% (24)	31% (36)
81–110 characters	104	Hi. I have an appointment for vaccination on Friday, but need to change the time. I do not find where I can do this	13% (14)	14% (15)	9% (9)	9% (9)	28% (29)	27% (28)

All example message sequences translated from Norwegian by the first author.

(median = 15; 25th percentile = 9; 75th percentile = 25). The significance of the difference was evident in a Mann-Whitney U -test ($Z = -17.6$, $p < .01$)

5.4.2. Impact of chatbot design on user dialogue style

The impact of chatbot design on user dialogue style was investigated by comparing the proportion of utilitarian vs. social style dialogues for users of the two groups of chatbot design. While only a small proportion of users did apply a social dialogue style, more did so when the chatbot was presented in the lower right-hand corner of the webpage (107 of 1894 dialogues; 8%) than when it was presented centre-stage of the municipality website landing page (25 of 1716 dialogues; 2%). The difference was found to be significant in a Chi-square test ($X^2 = 56.9$, $p < .01$).

6. Discussion

In the discussion section, we will first discuss our findings relative to the presented background and research question. We then discuss the implications of our findings for theory and practice, before we reflect on study limitations and point out needed future research.

6.1. Users' interactions with chatbots in the public sector

There has been a lack of knowledge on how users interact with public sector chatbots. By presenting an analysis of dialogue data from interactions between users and a chatbot implemented at municipality websites, the study provides important insight into how users approach such interactions, what can be expected in terms of chatbot responses, and how the user and the chatbot may mutually impact the relevance and helpfulness of the interaction.

6.1.1. How users approach interactions with public sector chatbots (RQ1)

In our explorations of RQ1, user input to chatbot dialogue was found to be characterized by brevity and utility-orientation. The dialogues between users and chatbots typically included only one message sequence, messages were typically brief, and the users' dialogue style was typically utility-oriented. Service

efficiency is an important driver for digital government initiatives (e.g., US Department of State, 2012; EC, 2022), and is also acknowledged as a key service value of digital public sector services in general (Twizeyimana & Andersson, 2019) and public sector AI services in particular (Fatima et al., 2022; Makasi et al., 2020). The user input to chatbot dialogue analyzed in our study is aligned with this ambition towards efficient service provision, as users' brief messages and utility-oriented dialogue style arguably suggests that they aim to achieve their goals with limited concern for other aspects of the conversational interface provided by the chatbot.

The brevity and utility-orientation of user input is also in line with previous general observations of chatbot user motivation and experience. Brandtzaeg and Følstad (2017), in a survey study of chatbot users, found pragmatic motivations—including efficiency and convenience—to be key motivators of chatbot use. Furthermore, in a continuation of this work, they found the pragmatic quality of a chatbot to be a key determinant of chatbot user experience (Følstad & Brandtzaeg, 2020). Foundational user experience theory (Hassenzahl, 2018) suggests that the user experience of interactive systems may typically be strengthened by careful consideration also of hedonic quality, that is, the system's ability for stimulating and evocative experience catering to the identity of the user. However, the highly utility-oriented user interactions with the chatbot may suggest that for users of a public sector chatbot, efficient goal achievement is key—potentially even more so than for support chatbots of other sectors. For example, Liao et al. (2020), found that a substantial proportion of users of a company-internal conversational solution applied a playful and social dialogue style in their interactions. Likewise, Følstad and Taylor (2021) found a substantial proportion of users of a customer service chatbot to take on a social interaction style. Possibly, when accessing rights-based services through a chatbot, as is the case in public sector service provision, efficiency and utilitarian orientation is even more important than for other chatbot application areas.

6.1.2. What can be expected in terms of chatbot responses? (RQ2)

From our explorations of chatbot response characteristics, in response to RQ2, we found that most message sequences entailed a relevant chatbot response, and most dialogues returned help that likely was used by the user. Adequate

levels of response relevance and helpfulness are key to a utility-oriented chatbot. In particular as a lack of relevant responses or helpful dialogue outcomes is likely to compromise any user expectation of efficient goal achievement. The levels of response relevance and helpfulness found in this study parallels those found in studies of chatbots for customer services (Følstad & Taylor, 2021).

At the same time, as acknowledged in linguistic theory (Sacks et al., 1974; Schegloff, 2000), interaction through dialogue is likely to include instances of breakdown and repair where the parties need to identify and resolve misunderstandings as they spontaneously appear during dialogues. In users' interaction with chatbots, such breakdowns typically are evidenced by false positives and negatives, where users receive irrelevant or fallback answers to their chatbot requests (Ashktorab et al., 2019). Previous work on chatbots in the public sector have suggested that a lack of support in the chatbot to help users formulate sufficiently precise problem formulation, may be a cause of breakdown (Verne et al., 2022). In the studied chatbot, repair is initiated by the chatbot expressing uncertainty and offering alternatives in cases of insufficient user intent prediction confidence, which may for example happen in cases where users' problem formulation does not correspond to the conceptual framework of the service provider. Foreseeing potential sources of breakdown during dialogues, and seeking to proactively initiate conversational repair, will be important to public sector chatbots. It is noteworthy that the proportion of uncertainty responses in the studied chatbot outweighs false positives and negatives. This finding suggests that chatbot-initiated repair through expressing uncertainty and offering alternatives is indeed feasible in a public sector chatbot.

6.1.3. The mutual contribution of users and chatbot on response relevance and dialogue outcome (RQ3 and RQ4)

While it is valuable to gain insight into the characteristics of user input and chatbot responses, it may be as important to understand how the characteristics of user input and chatbot design may impact each other. Hence, the third and fourth sub-questions of the study concerned the impact of the users' approach to the chatbot interaction on response relevance and dialogue outcome (RQ3), and the impact of chatbot design on users' approach to interaction (RQ4). In a dialogue system resembling that of conversation between people, interaction typically not predetermined but can vary for example in terms of content, dialogue length, and turn or sequence characteristics (Sacks et al., 1974). Here, the partners in the conversation adapt in response to each other's contributions, which is for example seen in conversational repair (Ashktorab et al., 2019). Ability for such adaptation of conversation is sought in conversational agents, such as public sector chatbots, e.g., by adhering to key principles of conversation in agent design (Moore, 2018).

Our findings lend support to the notion that user input and chatbot response impact each other mutually during interaction. We find it particularly intriguing that users' way of phrasing their requests may impact chatbot response relevance and helpfulness, while at the same time chatbot design may impact users' way of phrasing their requests. Specifically, we find that users who present their requests in

relatively brief messages are more likely to receive a relevant response and eventually reach a helpful outcome of the conversation. Such efficient input and response interactions are likely to be important to users' performance expectations for chatbots in the public sector, and potentially lead to strengthened chatbot uptake (Abbas et al., 2022). In particular, as such chatbots typically are intended to serve as support for service triaging (Makasi et al., 2022). At the same time, the mutuality in chatbot design and user's response brevity suggests that chatbot developers may have an opportunity to further strengthen efficiency in interaction by designing for interactions where users are nudged into providing input at an adequate level of detail.

We found a chatbot presented centre-stage of the municipality webpage to be associated with a tendency for briefer user input and a strengthening of response relevance and dialogue outcome compared to a chatbot presented on the lower right side of the municipality webpage. Possibly, this effect may be due to the different associations that different chatbot designs trigger in the users (Hall, 2018). A chatbot presented in the lower right side of the municipality webpage may possibly trigger associations of chat dialogues with human personnel, as manned customer service chats are typically made available through an interaction design involving activation through an avatar located in the lower right side of the page. Conversely, this association with manned customer service chat may be weaker for a chatbot presented centre-stage on the municipality landing page. Regardless of the underlying mechanism, this effect of variation in chatbot design on how users phrase their input and, subsequently, on the quality of the chatbot response, suggests the need to be mindful of the effect of design cues on user behaviour, as well as of the user behaviour desired for optimal chatbot performance. Potentially, as chatbots become increasingly advanced and capable to provide service negotiation rather than service triage (Makasi et al., 2022), it may be beneficial to design for other user behaviours. For example, in generative chatbots based on large language models, such as ChatGPT, there is a tendency towards user acceptance of longer user messages to allow for more detailed and precise chatbot responses (Skjuve et al., 2023). Hence, while our findings may be taken to mean that current chatbots for public sector services may benefit from encouraging brief user input, in the long term the main takeaway should be the need to consider how chatbot design impacts user behaviour, and which user behaviour will most likely lead to a desirable outcome.

6.2. Implications

The presented findings hold important implications for theory and practice. In the following, we have summarized what we consider to be the key implications.

6.2.1. Implications for theory

The key implications of the study findings for theory includes the following:

- *Alignment of user behaviour, theory on public value, and background on chatbot user experience:* Drawing from

previous research on public value (Makasi et al., 2020; Twizeyimana & Andersson, 2019) and user experience of chatbots in public sector (Abbas et al., 2022; Makasi et al., 2022), efficiency is key to service provision through chatbots. Our findings show that this aim for efficiency in service provision is also reflected in user behaviour. This consistency in user behaviour, user perceptions, and service provider aims is promising for future theory development in the field.

- *Mutual impact of users and chatbot on response relevance and dialogue outcome:* A noteworthy implication for theory in our findings is the dependency between users and chatbot with regards to response relevance and dialogue outcome. Dialogue design concerns facilitating desirable outcomes, though achieving such outcomes also may require an alignment with users' perceptions and behaviour (Hall, 2018). The mutual impact of users and chatbot identified in this study may motivate theory development to strengthen our understanding of this interplay.
- *Incremental improvement through efficiency gains:* Van Noordt and Misuraca (2019), in an analysis of chatbots in public sector services, argued that such chatbots seem more likely to provide incremental improvement rather than transformational change. Our findings corroborate this argument, and also expand on it as we show the importance of efficiency gains to achieve the foreseen incremental improvement by deploying chatbots in the public sector.

6.2.2. Implications for practice

Key implications for practice include the following three:

- *The prominence of efficiency in interaction:* For practical design and deployment of chatbots, our findings suggest that efficiency in interactions should have prominence. This may, in particular, be true for a chatbot for service triage (Makasi et al., 2022), as providing efficient leads to needed information and support is key for such chatbots.
- *Design to encourage desired user behaviour:* To achieve efficient interactions, the design and deployment of chatbots should not only be done in consideration of chatbot content and services, but also with concern for how chatbot design may impact user behaviour. Our findings suggest the importance of awareness on whether and how a given chatbot design may lead to desired user behaviour.
- *Monitor effects of design choices on user behaviour:* While some effects of design choices on user behaviour may be determined up front, through adherence to design guidelines, other effects may be more challenging to predict. Our findings identified an effect of chatbot design on user behaviour which has not been discussed in existing guidelines. Other such effects are likely to be identified in the future, and effects may also be due to contextual aspects of the service provision. Hence, public sector service providers deploying chatbots should monitor the effect of their design choices on user behaviour and, subsequently, on chatbot response relevance and helpfulness. This to identify successful design choices and improve on counter-productive choices.

6.3. Limitations and future research

The study is to our knowledge the first exploration of general user interactions with a public sector chatbot based on dialogue analysis. As such, the study provides a needed complement to previous work based on interview studies and questionnaires. At the same time, the study has important limitations which point forward to needed future research.

First, the study is conducted on user interactions with the single chatbot over a relatively short time window. While the chatbot is implemented in a range of municipalities with substantial variation in design, the core technology and content of chatbot nevertheless is common to all these implementations. We see this limitation as particularly important given the rapid development within the field of chatbot technology. Hence, the findings presented in this study should be replicated and expanded by research involving other chatbots. For example, chatbots based on other platforms or chatbots addressing other service levels in the taxonomy of Makasi et al. (2022).

Second, the study is conducted in one specific service context: Norwegian municipalities. While this service context shares strong resemblance with public sector service provision at other levels of government or in other countries, there is nevertheless a need to validate and complement the findings of this study with findings from studies at other levels of government and in other countries. In particular, we foresee studies including cross-sectorial and regional comparisons of user interactions with public sector chatbots.

Third, following from the anonymity of the chatbot users, we do not have information of demographic data. This limited the scope of the study, as we could not investigate possible effects of demographic variation on users' approaches to chatbot interaction or the relevance and helpfulness of chatbot output. Furthermore, the lack of demographic data disallowed checks for demographic bias in sampling. It will be valuable in future research to include users' demographic data to address this limitation in our study.

Fourth, the study is mono method, in the sense that only chatbot log data is considered. While this approach serves to complement existing research, it does not combine different data sources and methods to allow for triangulation. Future research including log data as well as data from users' self-reports through interviews or questionnaires would, hence, be a needed complement to this study. Here, it would be particularly relevant to investigate whether also other aspects of chatbot interaction than efficiency, such as a sense of sociability, may be important to the user experience. Also, we foresee observational studies to allow for in-depth understanding of the users' context and how this may impact experience and dialogue outcome.

Finally, the study is conducted at a single point in time with data collection spanning a few weeks. This disallows studies of change in user behaviour and dialogue outcomes. In future studies, we foresee the inclusion of multiple data collection points to allow for longitudinal analysis of change and understanding of long-term impact of chatbot design choices.

In spite of its limitations, the presented study provides an initial step towards needed knowledge on how users' interact with chatbots in the public sector. We trust that the findings will motivate future research in this important and interesting line of research.

Disclosure statement

The second author is an employee of Prokom, the company providing the chatbot of this study. This potential conflict of interest was mitigated by the study being led by an independent research organization, SINTEF, and by using an established approach to dialogue analysis.

Funding

This work was supported by the European Commission, Horizon 2020 Research and Innovation Action ETAPAS, grant no. 101004594.

References

- Abbas, N., Følstad, A., & Bjørkli, C. A. (2022). Chatbots as part of digital government service provision – a user perspective. In *Proceedings of CONVERSATIONS 2022 – 6th International Workshop on Chatbot Research and Design* (pp. 66–82). Springer International Publishing.
- Adiwardana, D., Luong, M. T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., & Le, Q. V. (2020). *Towards a human-like open-domain chatbot*. arXiv preprint arXiv:2001.09977
- Amiri, P., & Karahanna, E. (2022). Chatbot use cases in the Covid-19 public health response. *Journal of the American Medical Informatics Association: JAMIA*, 29(5), 1000–1010. <https://doi.org/10.1093/jamia/ocac014>
- Androutopoulou, 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. <https://doi.org/10.1016/j.giq.2018.10.001>
- Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, 37(4), 101490. <https://doi.org/10.1016/j.giq.2020.101490>
- Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. In *Proceedings of CHI 2019. Paper no. 254*. ACM.
- Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In *International conference on internet science* (pp. 377–392). Springer.
- Carter, L., & Bélanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal*, 15(1), 5–25. <https://doi.org/10.1111/j.1365-2575.2005.00183.x>
- European Commission (2020). *AI watch - artificial intelligence in public services. Technical report*. AI Watch, European Commission. <https://publications.jrc.ec.europa.eu/repository/handle/JRC120399>
- European Commission (2022). *New digital government policy: Conclusion paper*. European Commission. <https://digital-strategy.ec.europa.eu/en/library/new-digital-government-policy-conclusion-paper>
- Fatima, S., Desouza, K. C., Buck, C., & Fiel, E. (2022). Public AI canvas for AI-enabled public value: A design science approach. *Government Information Quarterly*, 39(4), 101722. <https://doi.org/10.1016/j.giq.2022.101722>
- Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, 5(1), 1–14. <https://doi.org/10.1007/s41233-020-00033-2>
- Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., von Wolf, R. M., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing*, 103(12), 2915–2942. <https://doi.org/10.1007/s00607-021-01016-7>
- Følstad, A., & Taylor, C. (2021). Investigating the user experience of customer service chatbot interaction: A framework for qualitative analysis of chatbot dialogues. *Quality and User Experience*, 6(1), 1–17. <https://doi.org/10.1007/s41233-021-00046-5>
- Hansson, S. O., & Fröding, B. (2022). *Code of conduct for disruptive technology in the public sector*. ETAPAS – Ethical Technology Adoption in Public Administration Services. <https://www.etapas-project.eu/resources/deliverables/>
- Hall, E. (2018). *Conversational design*. A Book Apart.
- Hassenzahl, M. (2018). The thing and I (Summer of '17 Remix). In M. Blythe & A. Monk (Eds.), *Funology 2: From usability to enjoyment* (pp. 17–31). Springer.
- Henk, A., & Nilssen, F. (2021). *Is AI ready to become a state servant? A case study of an intelligent chatbot implementation in a Scandinavian public service* [Paper presentation]. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)* (pp. 5515–5524), University of Hawaii, Manoa. <https://doi.org/10.24251/HICSS.2021.670>
- Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3), 221–236. <https://doi.org/10.1016/j.giq.2015.07.001>
- Liao, Q. V., Geyer, W., Muller, M., & Khazaen, Y. (2020). *Conversational interfaces for information search. Understanding and improving information search* (pp. 267–287). Springer.
- Makasi, T., Nili, A., Desouza, K. C., & Tate, M. (2022). A typology of chatbots in public service delivery. *IEEE Software*, 39(3), 58–66. <https://doi.org/10.1109/MS.2021.3073674>
- Makasi, T., Nili, A., Desouza, K., & Tate, M. (2020). Chatbot-mediated public service delivery: A public service value-based framework. *First Monday*, 25(12). <https://doi.org/10.5210/fm.v25i12.10598>
- Makasi, T., Nili, A., Desouza, K., & Tate, M. (2022). Public service values and chatbots in the public sector: Reconciling designer efforts and user expectations. In *Proceedings of the 55th Hawaii International Conference on System Sciences* (pp. 2334–2343), University of Hawaii, Manoa. <https://doi.org/10.24251/HICSS.2022.292>
- McTear, M. (2021). *Conversational AI: Dialogue systems, conversational agents, and chatbots*. Morgan & Claypool.
- Moore, R. J. (2018). A natural conversation framework for conversational UX design. In *Studies in conversational UX design* (pp. 181–204). Springer.
- Panagiotopoulos, P., Klievink, B., & Cordella, A. (2019). Public value creation in digital government. *Government Information Quarterly*, 36(4), 101421. <https://doi.org/10.1016/j.giq.2019.101421>
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1), 1–63. <https://doi.org/10.1017/S0047404500001019>
- 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 *Proceedings of the international conference on electronic participation* (pp. 133–146). Springer.
- Skjuve, M., Brandtzaeg, P. B., Følstad, A. (2023). *Why people use ChatGPT*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4376834
- Stahl, B. C. (2005). The ethical problem of framing e-government in terms of e-commerce. *The Electronic Journal of e-Government*, 3(2), 77–86.
- Tisland, I., Sodefjed, M. L., Vassilakopoulou, P., & Pappas, I. O. (2022). The role of quality, trust, and empowerment in explaining satisfaction and use of chatbots in e-government. In *Proceedings of the Conference on e-Business, e-Services and e-Society* (pp. 279–291). Springer.
- Twizeyimana, J. D., & Andersson, A. (2019). The public value of E-government—a literature review. *Government Information Quarterly*, 36(2), 167–178. <https://doi.org/10.1016/j.giq.2019.01.001>
- US Department of State (2012). *Digital government strategy*. <https://www.state.gov/digital-government-strategy/>
- Van Noordt, C., & Misuraca, G. (2019). New wine in old bottles: Chatbots in government. In *Proceedings of the international conference on electronic participation* (pp. 49–59). Springer.

- Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, I. O. (2023). Developing human/AI interactions for chat-based customer services: Lessons learned from the Norwegian government. *European Journal of Information Systems*, 32(1), 10–22. <https://doi.org/10.1080/0960085X.2022.2096490>
- Verne, G. B., Steinstø, T., Simonsen, L., & Bratteteig, T. (2022). How can I help you? A chatbot's answers to citizens' information needs. *Scandinavian Journal of Information Systems*, 34(2), 233–280.
- Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997). *PARADISE: A framework for evaluating spoken dialogue agents*. arXiv preprint [cmp-lg/9704004](https://arxiv.org/abs/cmp-lg/9704004)

About the authors

Asbjørn Følstad is chief scientist at SINTEF, a Norwegian independent research organization, with a PhD in psychology from the University of Oslo. His field of research is human-computer interaction. Key research interests are conversational user interfaces and human-centred AI.

Nina Bjerkreim-Hanssen is a senior AI-trainer at Prokom, and part of the team developing and maintaining the municipality chatbot Kommune-Kari. She has contributed to the presented research through the European Horizon 2020 research and innovation action ETAPAS.