Public Service Values and Chatbots in the Public Sector: Reconciling Designer efforts and User Expectations

Tendai Makasi
Queensland University of Technology, Brisbane
tendai.makasi@hdr.qut.edu.au

Kevin C. Desouza
Queensland University of Technology, Brisbane
kevin.desouza@qut.edu.au

Alireza Nili
Queensland University of Technology, Brisbane
a.nili@qut.edu.au

Mary Tate
Victoria University of Wellington, Wellington
mary.tate@vuw.ac.nz

Abstract

Chatbots are deployed across a wide range of public services, frequently to manage the increased volumes of online service requests. The appropriateness of many chatbot initiatives is often challenged. One reason for this is these initiatives are largely driven by agency centric goals, often neglecting the expectations of other public stakeholders. A public service value perspective – founded on the notion of public value – offers an avenue to represent the views of other public stakeholder groups. We examine the public service values of two key stakeholder groups – designers and users and discuss how they can be reconciled.

1. Introduction

Chatbots are computer programs that exploit Natural Language Processing (NLP) techniques to emulate human-human conversations during service interactions [1, 2]. They facilitate service interaction by generating either text-based or voice based responses to user enquiries using data collected from the current and previous inquiries [3]. Public agencies deploy chatbots as digital service agents across a range of services. The benefits of chatbots for public agencies include reduced service delivery costs and reduced employee workloads [1]. For service users, the use of chatbots to deliver reduced waiting times for service assistance, and faster access to domain-specific knowledge are the main benefits [2]. For example, in the fight against the COVID-19 pandemic, many health institutions across the globe [e.g. 4, 5] introduced chatbots to perform quick health assessments and provided critical health information to the public while minimizing the numbers of physical visits to the premises, or long waits in telephone queues [1].

Despite public agencies increasingly turning to chatbot-mediated service delivery channels, these initiatives often fall short of satisfying all the service delivery expectations [6]. Chatbots should exhibit appropriate public service values such as fairness. Those that do not, risk fierce criticism in the "court of public opinion". For example, RITA – a chatbot designed by the Transport for New South Wales in Australia was branded a waste of public funds after continuously failing to adapt to different user circumstances, which resulted in discriminative (or selective) service assistance [7]. Similarly, the National Health Services (NHS) – UK, faced backlash from both the citizens and the medical professionals over the introduction of an AI-powered Babylon chatbot technology [8]. The backlash led to a scaled down initiative which facilitates only the less invasive triaging service assistance (e.g. recommending patients to doctor consultations or over-the-counter help from a pharmacy)

Technologies are never value neutral [9]. The increase in technology adoption in public service delivery has resulted in calls to ensure creating public value is the ultimate goal of such initiatives [10]. Public service value describes the values held by public service agents during service design and delivery to ultimately ensure creation of public value [11]. Despite the public service value notion being well discussed within the public management domain, there remains little understanding of its intersection with technology-driven service delivery, particularly when emerging technologies such as chatbots are being used [10, 12]. Few studies have highlighted the need to consider chatbot specifications based on specific public service contexts [3, 6, 13]. Consequently, existing research literature offers limited understanding of how chatbots accommodate

public service value within the public sector. Chatbots used in public services are made up of a composition of design choices made by the chatbot designers based on specifications provided by the public agencies [14]. We know little about what value dimensions are prioritized by chatbot designers. There is also limited knowledge about the importance of these value dimensions to the chatbot users and what they expect. Paying insufficient attention to public service value dimensions can reduce the likelihood of acceptance by the public, and can lead to controversies [15].

To ensure chatbot initiatives within the public sector holistically contribute towards public service value, we investigated stakeholder groups directly involved in public sector chatbot initiatives including the chatbot users and chatbot designers. We address the question "What are the public service value dimensions that need to be considered for designing chatbots in chatbot-mediated public services (a) from the designer perspective and (b) from the user perspective?".

Considering the two perspectives enables understanding the impact of introducing chatbots on those directly involved in the design, development, implementation and use of chatbots. Specifically, we seek to reconcile these two public stakeholder group accounts of chatbot-mediated public service value for two reasons including (1) informing the design choices (choices of values which are driven partly by the requirements defined by the public agencies and other times by the individual choices of the chatbot designers) [16]; and (2) understanding the user perspective of value expectations from chatbot initiatives in public sector. Understanding and reconciling these viewpoints is critical towards designing chatbots that contribute towards the creation of public value holistically [17].

2. Study Background

2.1. Chatbots in public service delivery

Public agencies often deploy the task-oriented chatbots which are designed to assist people with specific service inquiries. These chatbots can be classified into specific types based on the technical features and capabilities [3]. Following [3, 18], we consider the algorithmic approach as the key distinguishing criteria for categorising chatbots. Chatbots employ either a rule-based approach or an AI/ machine learning driven approach. Rule-based chatbots exhibit basic functional capabilities. We refer

to these as basic chatbots. On the other hand, AI/machine learning driven chatbots exhibit advanced functional capabilities. We refer to these as the advanced chatbots.

Basic chatbots follow a retrieval based model where responses to user queries are linked to pre-determined responses in a manner similar to moving through the branches of a tree diagram [19]. Basic chatbots use a repository or knowledge base (e.g. FAQ responses database) of predefined responses with a ranking model that selects an appropriate response matching the user's enquiry [20]. Basic chatbots can either lead highly structured interactions (e.g. MANDI[1], TfL TravelBot[2]), or can handle evolving semi-structured interactions (e.g. CHIP[3], RAMMAS[4]). Basic chatbots can interpret individual requests and match these with the linked data sources to provide the appropriate information and service links for the individual.

Advanced chatbots follow a generative-based model, leveraging machine learning capabilities to exploit a knowledge base of multiple repositories to generate responses to user queries. The algorithms that generate responses improve overtime by automatically learning user inquiry patterns [19]. Advanced chatbots can retain interactional data to create user profiles, and have capabilities to keep track of context, thus facilitate evolving interactions with the users. Further, advanced chatbots assist with generating potential service recommendations based on the rich user profiles [21], and sometimes for negotiating personalized service options with the users [e.g. 22].

2.2. Public service value in chatbot mediated service delivery

Public services are often complex and need to cover diverse citizens' needs and circumstances [23]. Public agencies may struggle with institutional inertia and often tend to be risk averse due to reputational risk related to appropriate use of public funds (e.g., investing in chatbot-mediated service delivery should create value for service applicants) [24]. Public value theory provides a narrative that can help to explain the contributions of digital technologies such as chatbots in the public sector using broader criteria than merely cost effectiveness [10]. Chatbots serve as digital agents during service delivery, often taking the place of human agents. Public service values describes the values held by public agents during service delivery to facilitate creating public value [6, 11]. Deploying chatbots for public service delivery requires attention

---

to the notion of public service value to ensure a range of benefits for the public, including better and more flexible services, improved effectiveness of service delivery, and improved sense of safety and trust in government [6, 11].

There is limited empirical research that focuses on chatbot-mediated public service value dimensions. Chatbot design research tends to focus indirectly on enhancing only a subset of the chatbot-mediated public service value dimensions. A literature review of current chatbot design research towards facilitating public service value dimensions described in [6] shows that user orientation receives the highest levels of research attention. Recommendations for enhancing user orientation include (1) developing chatbots with emotional intelligence [25]; (2) incorporating 'direct address' ("i.e., *including the user's name in an utterance directed to the user*") in chatbot interactions [26]; (3) incorporating additional and advanced APIs (e.g. Google Translate, Google Maps) in chatbots to orient the chatbot conversational capabilities towards better understanding the user's requirements [25]; and (4) developing chatbot interactive capabilities that facilitate incremental service negotiations that go beyond a single interaction experience [27].

Other studies have also investigated different chatbot designs to indirectly address other public service value dimensions. For example, research towards integrating multimodal chatbot interfaces can enhance the four public service value dimensions including effectiveness [28], adaptability [29] and collaborative intelligence [28]. Enrichening chatbot responses with social signals has been studied as a means to enhance trust [29]. Considering professionalism, the importance of chatbot context awareness with regards to cultural sensitivity has been examined [30]. The degree to which direct addressing during chatbot-human interaction affects perceptions of professionalism was investigated in an experimental study [26]. We contribute to this discourse by comparing and reconciling the values of designers and users, within an overall public service value framework.

3. Methodology

Our approach involved investigations from two distinct perspectives: (1) user focused investigations and (2) designer focused investigations. The user focused investigations followed a laddering interviews approach to understand the specific public service value dimensions that contribute towards the users' decision to use (or not to use) a chatbot. The laddering interviewing technique enables soliciting relationships between specific chatbot attributes and value dimensions from chatbot users. The designer focused investigations followed a semi-structured interview approach to understand the public service value perspective of chatbot designers during their design work. Semi-structured interviews are suited to draw information from chatbot designers with an existing appreciation of how specific chatbot attributes influence the associated value dimensions. We triangulated the findings from the two participant groups to ensure coverage and reliability of the overall research findings.

3.1. Chatbot-user focused investigations

We adopted the A-C-V laddering interview technique. Laddering connects the observable attributes (A) of a product or service (e.g., technology affordances) with the perceived consequences, that is, the benefits or feelings experienced by the user (C). These in turn are linked to core values or beliefs held by the user (V) [36, 37]. Laddering is an interviewing approach that enables understanding how users translate the attributes of an artifact into meaningful value interpretations [31]. Laddering technique involves a series of directed probes to sequentially reveal the users' motives for choosing a particular service (or product) [32]. Laddering uses a hierarchical structure starting with user perception of the specific features of service (or product) and building from those to draw out the root reason of a user's decision about the product or service.

Laddering technique was adopted to understand the specific factors that motivate citizens (service consumers) to use chatbots to access public services. "*Laddering works best when respondents are providing associations while thinking of a realistic occasion in which they would use the product*" [31]. We used 3 scenario descriptions to ensure that all participants reflected on a realistic experience of chatbots within a similar context. With a chatbot presented as an optional service channel, the 3 scenario descriptions deliberately focused on three different service sectors with varying degrees of service complexity to capture a broad range of public services including : (1) a situation when the user unexpectedly loses their job and needs information relating to accessing government unemployment funding; (2) a situation when the user requires guidance with completing their financial income records for tax purposes; and (3) a situation when the user needs to make a specialist medical appointment. Participants were given time to familiarize themselves with each of the three scenario descriptions before the interviewing commenced.

Since the laddering technique enables linking the specific attributes of an artefact with value interpretations, we advertised for potential participants who had first-hand experience with a chatbot-mediated public service in Australia through social media posts on Facebook and LinkedIn. We conducted a total of 25 interviews with chatbot users, in addition to 3 pilot interviews that were initially conducted to refine the interview protocol. All the interviews were conducted over a period of 4-5 months. On average, each interview lasted between 20 to 40 minutes.

Through the laddering-based interview questioning, the study first solicited from each interviewee: (1) the desirable chatbot affordance (A) (higher-level enablement [33]) that influences the choice to use (or not use) a chatbot; (2) the consequence (C) of having the specified chatbot affordance during chatbot-mediated public service interaction; and (3) the core value (V) as the root value motivating the decision to use a chatbot to access the specific public service. In the case where the interviewee opted against using a chatbot, the interviewing followed an approach referred to as negative laddering – a widely acknowledged way of exploring hidden assumptions by formulating opposite meanings of the responses and probing for implications of these opposite meanings [31]. Table 1 shows two interview extracts comparing the two possible laddering interview questioning sequences.

Table 1: Examples of questions and user comments from the laddering interviews

Laddering interview (sample responses)	Negative laddering (sample responses)
Q: Would you use a chatbot in this scenario? A: "*Yes, I would*"	Q: Would you use a chatbot in this scenario? A: "*No, I wouldn't*"
Q: What is the main reason that would drive you to use a chatbot? A: "*availability of chatbot is 24/7 ... I am not worried about business hours*"	Q: What is the main reason that would make you opt not to use a chatbot? A: "*this is related to a very personalized issue ... a chatbot would take a lot of time*"
Q: Why is having someone to talk to when you need to important to you? A: "*I have had bad experience with call centres ... put in long queues, having to wait*"	Q: Why is not having 'a personalized experience' important to you? A: "*It's necessary to understand the intensity of the program required for me*"
Q: Why is avoiding long queues important to you? A: "*... the efficiency of getting help when I need it most.*"	Q: Why is being understood accurately important to you? A: "*It reduces the chances of incorrect recommendations*"

All data collected through the laddering interviews was analysed following the three steps suggested by Reynolds and Gutman [31] including content analysis, construction of an implication matrix, and construction of the hierarchical value map (HVM). Through content analysis, constructs were consolidated based on participant responses with similar underlying ideas. To establish connections between the constructs, all constructs were categorized into appropriate classes from affordances, consequences and then values. Next, for each of the laddering responses, the corresponding A-C-V sequence was established. All the A-C-V sequences were summarized to create an implication matrix that indicates the number of times each construct is linked directly (without any intermediary constructs) and indirectly (with at least one intermediary construct) to all the other constructs. To visualize the relationships between constructs, we generated a HVM. The widths of the interconnecting lines of the HVM show the strengths of the relationships between the constructs.

3.2. Chatbot-designer focused investigations

Semi-structured interviews are well suited to draw the individual thoughts of the participants using follow-up questions [34]. Guided by the chatbot-mediated public service value dimensions described in [6], we conducted semi-structured interviews with chatbot designers to explore the chatbot-mediated public service value dimensions from the chatbot designers' perspective; and determine how these value dimensions are accounted for during chatbot design.

The first section of the interviews aimed to profile the interviewees' chatbot design experience in terms of the number of years in chatbot design for public sector and the diversity of sectors. The second and third sections aimed to understand which public service value dimensions were prioritized during the design initiatives and included questions such as: what were the most important dimensions during the chatbot design, and how were these dimensions accounted for during chatbot design? The second section solicited responses based on the interviewees' previous chatbot design experiences. With the scarcity of advanced types of chatbots in the public sector [3], we envisioned that the chatbot designers' responses would most likely reflect insights that relate to the basic types of chatbots. To draw insights specific to advanced chatbots, we adopted a scenario-based approach that described requirements to design an advanced chatbot for use in the public sector in the third section. Using the scenario approach in research

provides a way to enhance preparedness and anticipate consequences of certain actions [35].

We identified potential participants (i.e., chatbot designers) with experience designing chatbots used in any specific public sector only in the commonwealth countries through their online profiles shared on LinkedIn. Invitations for participation in the interviews were sent individually to each of the potential participants. In addition, we used the snowballing technique (where each of the agreed participants suggested other potential chatbot designers) to identify more participants. In total, we conducted 12 interviews (including 3 pilot interviews to validate the interview protocol) with chatbot designers online via Zoom. The interviews were conducted over a period of 3-4 months. Each interview lasted on average between 40 minutes to 60 minutes.

The interview transcripts were analysed following a directed content analysis approach described by Hsieh and Shannon [36]. We established the initial codes of the content analysis as the 14 public service value dimensions. All transcripts were read, and relevant information was coded onto the respective codes (i.e., the value dimensions). Finally, all coded information was reviewed to understand how each public service value dimension was prioritized and accounted for during chatbot design.

4. Findings

We present the findings from the laddering interviews with the chatbot users (Section 4.1) and the findings from the semi-structured interviews with the chatbot designers (Section 4.2).

4.1. Chatbot users' perception of chatbot-mediated public service value dimensions

The initial step during the laddering interviews solicited the specific chatbot affordances that drew the interviewees towards using a chatbot. In total, we discovered eight affordances. Starting with the most popular affordance, these include: (1) ease of use – chatbot's ability to facilitate an interaction with minimum effort from the user; (2) predictability – chatbot's ability to provide the expected level of assistance for the user; (3) personalization – chatbot's ability to provide access to information that is relevant and appropriate to each user's specific circumstance; (4) reasoning competency – chatbot's ability to process the information shared by the user exhibiting acceptable levels of attention to detail; (5) quick responses – chatbot's ability to provide quick responses; (6) multi-modality – chatbot's ability to take into consideration information conveyed by the user in a non-verbal or non-textual form; (7) active links – chatbot's ability to provide links to relevant information; and (8) understanding capacity – chatbot's ability to accurately translate the information shared by the user to determine the exact user's needs.

We identified a total of six consequences linked to the chatbot affordances. Starting with the most frequent consequence, these include: (1) high accuracy – receiving accurate information; (2) no waiting – timely access to service assistance; (3) saves resources – minimized cost to access information; (4) information accessibility – facilitating easy access to relevant information; (5) information usability – receiving information in a readily usable format; and (6) clarification – receiving explanation of outcomes.

We identified a total of four value dimensions including (1) efficiency – achieving the appropriate outcome in a timely manner; (2) effectiveness – achieving the intended outcome with the information provided; (3) sustainability – maintaining an overall state of wellness; and (4) trust - having confidence in the accuracy of the outcome.

We generated a total of 75 A-C-V sequences. A total of 53 unique A-C-V sequences showed strong relationships between certain combinations of constructs. For example, 'ease of use' and 'efficiency' (10 direct links and 4 indirect link); and 'high accuracy' and 'effectiveness' (15 direct links) were the strongest interlinks. Figure 1 visualizes the strengths of the resulting relationships between the constructs.
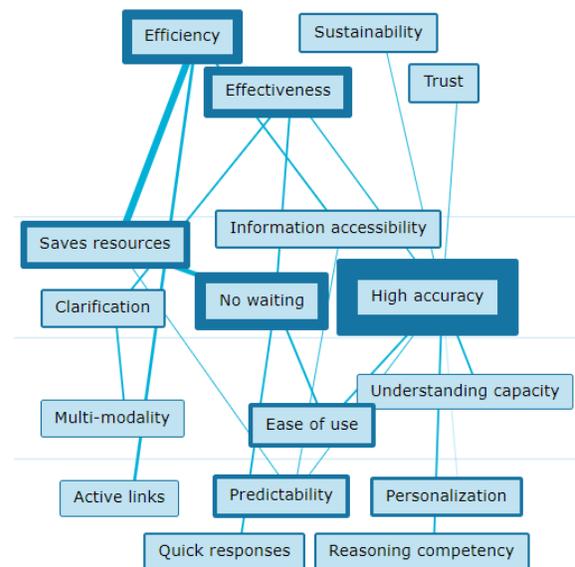
Figure 1: Hierarchical value map of chatbot affordances, consequences and relative value dimensions

4.2. Chatbot designers' perception of chatbot mediated public service value dimensions

First, we observed 3 distinct levels of how chatbot designers prioritize chatbot-mediated public service value dimensions during chatbot design efforts. The first level comprises of six dimensions discussed as high priority for both basic and advanced chatbots in the public sector. These include efficiency, effectiveness, user-orientation, professionalism, adaptability and trust in government. The second level comprises of five value dimensions cited as critical dimensions particularly in relation to chatbots that specifically retain user profiles (typically the advanced type of chatbots). These include privacy, legitimacy, accountability, acceptability and social license. The third level comprises of four dimensions acknowledged as highly important in terms of the overall performance management of all chatbot initiatives in the public sector. These include openness, sustainability, fairness and collaborative intelligence.

The importance of efficiency is expressed in two different perspectives. Efficiency is discussed in terms of optimizing the costs associated with chatbot initiatives in the public sector: *"you should make sure that the cost of having a (chat)bot is cheaper than the cost of having a human worker ... every feature on the (chat)bot* [should be] *cost effective"* – [Interviewee 1]. Also, designers discussed efficiency from the perspective of individual chatbot user leaning towards an improved and timely service experience: *"you have to consider how quickly and efficiently you can give feedback to the* [chatbot user] *because that will be the determinant of a successful user experience for them."* – [Interviewee 7].

Effectiveness is expressed in terms of how a chatbot provides accurate and relevant information to the user during the interaction: *"I always make sure the (chat)bot is working correctly, there's no bugs, it matches the answers correctly"* – [Interviewee 1]. A close interdependency between efficiency and effectiveness is emphasized by various interviewees e.g., *"effectiveness is like an outcome of an efficient design ... if the service is efficient, then the service is also effective and those go hand in hand for the user to essentially enjoy the experience"* – [Interviewee 4].

User-orientation is discussed in relation to how the chatbot is designed to best understand the user requirements during interaction: *"User orientation* [requires] *understanding who the chatbot is speaking to, making sure that the experience is correctly orientated to the user"* – [Interviewee 12].

Professionalism is discussed as important towards ensuring the users are not put off the main goal of interaction by unnecessary utterances from the chatbot: *"the chatbot should have a direct way of responding. It should not be sarcastic, but it should* [show] *a soft manner that treats everyone in a proper way."* – [Interviewee 7].

Adaptability is what ensures the chatbot remains a reliable source of service assistance in all circumstances. The designers acknowledged the importance of chatbots sensitivity to changing conditions such as updates in relation to information on service applicant's eligibility for a range of services. in the way public services are provided: *"we've seen with this COVID outbreak content we needed to update daily or maybe hourly regarding cancellations and [many other] things."* – [Interviewee 2].

The importance of privacy was discussed around protecting the identifiable components of individuals data collected during the interaction. Interviewees emphasized the need to avoid retaining and using sensitive user information unless it is absolutely necessary to do so: *"it's mostly about the storage of the conversations. If you build a user profile, what do you use it for and how you store it? ... conversation should be stored for a maximum time only to analyze if something went wrong, and access to those conversations should be limited ..."* – [interviewee 3].

Trust in government is influenced directly by the chatbot (during interaction): *"the one thing that the bot can do to ensure trust in the government is to provide accurate replies ... and making sure that the information it provides, and replies is up to date."* – [Interviewee 4]; and indirectly by the organization (beyond the interaction). The organization plays a role towards building trust in government by ensuring the users' information is kept private: *"People build trust when they know data is private and I imagine in the chatbot will be receiving fairly confidential and private information so it's really up to the organization to ensure those protocols are in place so that the user is confident that the data is being handled correctly."* – [Interviewee 6].

Legitimacy is acknowledged when chatbots are used to provide highly personalized service assistance that require chatbots to make use of the user's personal information. Many interviewees highlighted the interdependencies between legitimacy and both privacy and trust in government. Chatbots that collect and use personal user information should abide by the laws that govern the use and protection of personal information. In addition to providing accurate information, legitimacy is also upheld when the chatbot provides only responses that are both lawful

(compliant with the laws and regulations) and ethical. "*you've got to ensure that the chatbot responses [are] trustworthy and compliant to the laws and regulations ... If it gives a response that is unlawful, then it's totally going to destroy the credibility of the design and the solution.*" – [Interviewee 12].

The designers also discussed awareness of the importance of fairness for chatbots in the public sector. Designers acknowledged the need to extensively test the chatbots to eliminate any bias tendencies that affect fairness: "*Fairness has to be considered at the initial design stage and at the actual conversational (...) when a chatbot [exhibits] any form of discrimination, no one will use that bot.*"
" – [Interviewee 6].

Openness was discussed as another important factor. For example, one designer highlighted that "[*openness] is important to give confidence of privacy to the users*" – [Interviewee 10]. Another designer highlights that "*in government [openness] is key ... [it] is how the chatbot makes a decision, it's something that I believe research hasn't come to an answer yet.*" – [Interviewee 3].

Accountability was acknowledged as the degree to which the chatbot facilitates an accountable channel for service interaction. Further, it was discussed based on two factors including: (1) encouraging the users to share the most accurate information in order to receive an appropriate response; and (2) creating channels that enable the user to challenge an outcome that may not have been intended by the user: "*There are two sides, we need to make sure the chatbot provides information that can be verified as accurate and, two, the users have to give proper information so that the chatbot can provide accurate information*" – [Interviewee 11].

The importance of social license is discussed oriented towards ensuring an on-going approval of the chatbot initiative from the society at large. Steps have to be taken to communicate the benefits of introducing a chatbot in order for the society to embrace it: "*we aim for a solution that's providing some social good and people see it as a positive in terms of replacing an existing process. The biggest thing is educating users that this is now the [channel] they have to use to engage with the [organization]*" – [Interviewee 8].

In terms of acceptability, designers explained that in order to implement a successful chatbot in the public sector, the end users first have to accept the chatbot's viability as a service channel: "*users are the ones that should accept your end product at the end of the day... I always stick with the acceptability that is coming from the [user] and not from the developers' side*" – [Interviewee 2]. The interviewees also emphasized the importance of tracking the relevant metrics that give an indication of how the chatbot is received on a continuous basis.

Finally, sustainability was discussed in terms of ensuring that introducing chatbots is beneficial to the public sphere and particularly to the chatbot users. Introducing chatbots should ensure efficient utilization of public resources: "*government [agencies] ask for chatbots to help people and also to preserve money ... Implementing chatbots should reduce the pressure on human resources and financial resources for the chatbot users*". [Interviewee 11]

5. Discussion

We reconciled insights from the chatbot designers and chatbot users to (1) broaden the theoretical understanding of chatbot-mediated public service value; and (2) contribute towards the interpretation of public service value dimensions in the design of chatbot initiatives within the public sector.

Our findings lay both theoretical and practical foundations for advancing public-service value laden approaches in public sector chatbot initiatives. First, we observed that the chatbot users are primarily concerned with only a subset of chatbot-mediated public service value dimensions, specifically efficiency, effectiveness, sustainability and trust. On the other hand, chatbot designers acknowledge the importance of an additional 11 chatbot-mediated public service value dimensions (including user-orientation, professionalism, adaptability, privacy, legitimacy, accountability, acceptability, social license, openness, fairness and collaborative intelligence), however with different levels of priority. One reason for this disparity is that chatbot users focus primarily on their individual service experience, while chatbot designers are increasingly aware of the importance of public service values and embedding them in their designs. They mentioned that they are aware of the debates related to the notion of public service values. Many of them mentioned that they are observing more emphasis on values from public agencies commissioning the chatbot initiatives.

While some of the same values were important to both stakeholder groups, the detailed understanding of the values and how they should be measured were different. This suggests that both user perspectives and agency perspectives should be measured and reconciled. Additionally, specific chatbot design recommendations should envision a holistic interpretation of public service value dimensions to balance the technical aspects, functional aspects, and user experience.

Oftentimes public service value measures are defined from the viewpoint of the public agencies. For

example, while chatbot designers often discuss 'efficiency' in terms of how a chatbot contributes overall towards lowering the cost-of-service delivery for the public agencies, chatbot users interpret 'efficiency' from the viewpoint of achieving a service outcome in a timely manner. Consequently, to enhance efficiency, chatbot designers focus attention on developing chatbot algorithms that generate responses in order of priority based on frequently requested information. In contrast, chatbot users emphasize on facilitation of seamless service transition from chatbot to the relevant service as major contributing factors towards efficiency.

In terms of effectiveness, chatbot designers mainly focus on eliminating errors that may lead to incorrect chatbot responses while for chatbot users, effectiveness relates more to achieving the desired outcome based on their life circumstances and service needs. Chatbot designers also emphasize the need to implement intelligent dashboards for tracking the chatbot's performance to identify and eliminate algorithmic bugs. On the other hand, chatbot users feel the need for chatbots to afford them control over the interaction to accurately express their service needs as more important towards improved effectiveness. In addition to defining NLP databases of keywords, entities and intents aligned with expected user request to support the use of day-to-day vocabulary, chatbot designers also need to consider complementary features (e.g., clickable buttons and type-in options) for chatbot users to clarify service needs.

Trust in government established through trust in chatbot-mediated public service delivery is largely derived from the chatbot performance from the viewpoint of the chatbot users. Chatbot users are more concerned about the confidence they have in the accuracy of the final outcome and the levels of control over the information shared, particularly when confidential information is disclosed. The chatbot designers' envision trust is established when chatbots deliver outcomes that are comparable to other public service channels. Chatbot designers emphasize the need to minimize back-end access to user information collected during the interaction, unless it is necessary for algorithmic training purposes. All identifiable data elements are excluded for training. In cases where confidential user information is retained, chatbot designers also discuss the necessity of affording the users control to verify and modify such data.

Ensuring sustainability through chatbot-mediated public service delivery focuses more on the benefits generated towards the chatbot users. Chatbots should be designed to minimize the related service costs for the chatbot users. While the chatbot designers fail to explicitly provide chatbot features that can directly contribute towards sustainability, chatbot users emphasize that establishing sustainability requires minimizing complications towards achieving the intended outcomes.

Chatbot-mediated public service value dimensions do not exist in isolation. Establishing public service value dimension-specific chatbot design recommendations requires evaluating these recommendations against all the other public service value dimensions. For instance, many chatbot users highlight that chatbots fail to afford the opportunity to naturally communicate and negotiate specific service needs in detail. On the other hand, many chatbot research efforts are investigating how public service value dimensions including user-orientation and collaborative intelligence can be enhanced by introducing multi-modal chatbot interfaces. While research is emerging addressing the fallings of chatbots from the user perspective, care should be taken on the interplay of multi-modal chatbot interfaces particularly with the other public service value dimensions including privacy, openness and legitimacy. There is a risk that multi-modal chatbot interfaces will learn sensitive information during the interaction without the user's consent and knowledge. With the many public sector data governing bodies (e.g., GDPR) that regulate the use of citizens data in terms of both privacy and openness of procedures, there is an evident need for establishing legitimacy through compliance with existing laws.

Building social license is an ongoing expectation of public service delivery. The performance of chatbots deployed in the public sector over time and across all the public service value dimensions determines the ongoing social license and public trust in chatbot-mediated public service delivery. From the interviews, particularly the interviews with the chatbot designers, we realized that both social license and trust in government are closely influenced by professionalism, adaptability, accountability, acceptability and fairness. During chatbot-mediated service interactions, it is necessary to ensure that chatbots carefully consider the choice of words and maintain a respectful tone to uphold professionalism. Similarly, given the ease of access of chatbots to users, it is essential that the chatbots display adaptability in consistently delivering accurate and up to date information. In terms of accountability, chatbots should strive to support traceable decision-making process records. In public sector chatbot initiatives, acceptability can be accounted for when platforms to engage diverse public stakeholder groups are established throughout the chatbot design process. With regards to fairness, it is essential to invest in rigorous chatbot testing with expected bias inputs to

ensure the chatbot responses are free from any forms of bias.

6. Future Research and Limitations

First, future research can use findings of our empirical research (public service value dimensions for chatbot-mediated public service delivery) as an initial guide for developing and proposing design guidelines and recommendations that support embedding the values in the design of chatbots. Overall, public service value dimensions play a role towards the user acceptance of digital systems in the public sector [37], we invite future research to investigate this notion in the context of chatbots.

Our empirical research is limited to two different types of techniques to interview two major groups of stakeholders (designers and users). While our study is one of the few of its kind that uses more than one method of information elicitation and covers insights from more than one group of stakeholders, we believe that future researchers can contribute to a higher coverage of potentially new insights by conducting different methods (e.g. focus groups and surveys) to collect data from the chatbot users, chatbot designers and potentially different stakeholder groups such as public managers. The survey method could also be used to quantitatively test our research findings. Our user focused investigations were limited to individuals with chatbot-mediated service experience within an Australian public service context. Future research can also explore how different public (and private) service contexts and user demographics, including inexperienced chatbot users may influence the findings.

Our research was conducted in the midst of the COVID-19 pandemic, where a number of individuals turned towards digital service channels [38]. We acknowledge that this might have influenced our findings. Given the inexistence of previous studies for comparison in this research area, we invite future research to investigate whether pandemics such as COVID-19 may impact the value perceptions of users and designers of digital systems in the public sector.

7. Conclusion

Embracing public service value dimensions at the centre of public sector chatbot initiatives is critical to align these initiatives to contribute towards public value. There is increased risk of drawing negative attention on chatbot initiatives that overlook the public service value dimensions. Chatbot users and chatbot designers are two key stakeholder groups that influence the success of chatbot initiatives in the public sector. Our paper presents insights to incorporate the value perspectives of these two key chatbot-mediated public stakeholder groups. In addition, the insights from this paper can be used to propel future research studies on integrating public service value theories with AI driven service delivery concepts.

8. References

[1] A. S. Miner, L. Laranjo, and A. B. Kocaballi, "Chatbots in the fight against the COVID-19 pandemic," *npj Digital Medicine,* vol. 3, p. 65, 2020/05/04 2020.

[2] A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis, "Transforming the communication between citizens and government through AI-guided chatbots," *Government Information Quarterly,* vol. 36, pp. 358-367, April 2019.

[3] T. Makasi, A. Nili, K. C. Desouza, and M. Tate, "A Typology of Chatbots in Public Service Delivery," *IEEE Software,* 2021.

[4] Centers for Disease Control and Prevention, "COVID-19 Testing Overview," ed: U.S. Department of Health & Human Services, 2020.

[5] SingHealth, "SingHealth and A*STAR co-develop smart chatbot to enhance care for COVID-19 patients at community care facilities," ed: SingHealth Group, 2020.

[6] T. Makasi, A. Nili, K. C. Desouza, and M. Tate, "Chatbot-mediated public service delivery," *First Monday,* 2020.

[7] B. Graham, "Transport for NSW's interactive Facebook assistant slammed for posting error messages and website links," ed: news.com.au, 2018.

[8] D. Finley. (2020, 27 February) Virtual care firm Babylon fired back at a doctor critiquing its chatbot by publicly posting his data on Twitter. *Business Insider.* Available: https://images.markets.businessinsider.com/babylon-chatbot-fires-back-at-doctor-critic-on-twitter-2020-2

[9] P. Vermaas, P. Kroes, I. van de Poel, M. Franssen, and W. Houkes, "A philosophy of technology," in *From technical artefacts to sociotechnical systems* vol. 6, ed: Morgan & Claypool, 2011, pp. 1-134.

[10] P. Panagiotopoulos, B. Klievink, and A. Cordella, "Public value creation in digital government," ed: Elsevier, 2019.

[11] E. Witesman and L. Walters, "Public service values: A new approach to the study of motivation in the public sphere," *Public Administration,* vol. 92, pp. 375-405, 2014.

[12] L. Andrews, "Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'," *Public Administration,* vol. 97, pp. 296-310, 2018.

[13] I. Cantador, J. Viejo-Tardío, M. E. Cortés-Cediel, and M. P. Rodríguez Bolívar, "A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government," in *DG. O2021: The 22nd Annual International Conference on Digital Government Research*, 2021, pp. 168-179.

[14] D. D. Luxton, "Ethical implications of conversational agents in global public health," *Bulletin of the World Health Organization,* vol. 98, p. 285, 2020.

[15] J. van Doorn, M. Mende, S. M. Noble, J. Hulland, A. L. Ostrom, D. Grewal*, et al.*, "Domo Arigato Mr. Roboto:Emergence of Automated Social Presence in Organizational Frontlines and Customers' Service Experiences," *Journal of Service Research,* vol. 20, pp. 43-58, 2017.

[16] J. Van den Hoven, "Value sensitive design and responsible innovation," *Responsible innovation: Managing the responsible emergence of science innovation in society,* vol. 47, pp. 75-83, 2013.

[17] B. Friedman, "Value-sensitive design," *interactions,* vol. 3, pp. 16-23, 1996.

[18] M. Riikkinen, H. Saarijärvi, P. Sarlin, and I. Lähteenmäki, "Using artificial intelligence to create value in insurance," *International Journal of Bank Marketing,* 2018.

[19] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir*, et al.*, "Conversational agents in healthcare: a systematic review," *Journal of the American Medical Informatics Association,* vol. 25, pp. 1248-1258, 2018.

[20] J. L. Z. Montenegro, C. A. da Costa, and R. R. da Rosa, "Survey of conversational agents in health," *Expert Systems with Applications,* vol. 129, pp. 56-67, 1 September 2019.

[21] H. Crouch, "GP at Hand expansion reduced after NHS England objections," ed: Digital Health, 2018.

[22] J. Hendry. (2018, 10 December) NDIS' great bot hope Nadia takes more time off for stress leave. *IT News*. Available: https://www.itnews.com.au/news/ndis-great-bot-hope-nadia-takes-more-time-off-for-stress-leave-516592

[23] F. Bannister and R. Connolly, "ICT, public values and transformative government: A framework and programme for research," *Government Information Quarterly,* vol. 31, pp. 119-128, 2014/01/01/ 2014.

[24] A. Nili, A. Barros, and M. Tate, " The public sector can teach us a lot about digitizing customer service," *MIT Sloan Management Review,* vol. 60, pp. 84-87, 2019.

[25] K. Ralston, Y. Chen, H. Isah, and F. Zulkernine, "A Voice Interactive Multilingual Student Support System using IBM Watson," *arXiv preprint arXiv:.00471,* 2019.

[26] H. Candello, C. Pinhanez, M. Pichiliani, M. Vasconcelos, and H. Conde, "Can direct address affect user engagement with chatbots embodied in physical spaces?," in *Proceedings of the 1st International Conference on Conversational User Interfaces*, 2019, pp. 1-9.

[27] A. Rosenfeld, I. Zuckerman, E. Segal-Halevi, O. Drein, and S. Kraus, "NegoChat: a chat-based negotiation agent," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 525-532.

[28] M. R. Ali, D. Crasta, L. Jin, A. Baretto, J. Pachter, R. D. Rogge*, et al.*, "LISSA—Live Interactive Social Skill Assistance," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 173-179.

[29] Y. Sakai, Y. Nonaka, K. Yasuda, and Y. I. Nakano, "Listener agent for elderly people with dementia," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 199-200.

[30] M. Makatchev, I. Fanaswala, A. Abdulsalam, B. Browning, W. Ghazzawi, M. Sakr*, et al.*, "Dialogue patterns of an arabic robot receptionist," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010, pp. 167-168.

[31] T. J. Reynolds and J. Gutman, "Laddering theory, method, analysis, and interpretation," *Journal of advertising research,* vol. 28, pp. 11-31, 1988.

[32] C. G. Russell, A. Busson, I. Flight, J. Bryan, J. V. L. Van Pabst, and D. N. Cox, "A comparison of three laddering techniques applied to an example of a complex food choice," *Food quality preference,* vol. 15, pp. 569-583, 2004.

[33] A. Nili, M. Tate, and A. Barros, "A disciplined approach for enhancing the technology acceptance model," in *European Conference on Information Systems (ECIS) 2020 Proceedings*, 2020.

[34] K. E. Newcomer, H. P. Hatry, and J. S. Wholey, "Conducting semi-structured interviews," in *Handbook of practical program evaluation.* vol. 492, ed, 2015.

[35] K. C. Desouza, "Scenario management from reactivity to proactivity," *IT Professional,* vol. 7, pp. 42-48, 2005.

[36] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qualitative health research,* vol. 15, pp. 1277-1288, 2005.

[37] S. R. Chohan and G. Hu, "Success factors influencing citizens' adoption of IoT service orchestration for public value creation in smart government," *IEEE Access,* vol. 8, pp. 208427-208448, 2020.

[38] D. Agostino, M. Arnaboldi, and M. D. Lema, "New development: COVID-19 as an accelerator of digital transformation in public service delivery," *Public Money Management Research Review,* vol. 41, pp. 69-72, 2021.