What do citizens think of AI adoption in public services? Exploratory research on citizen attitudes through a social contract lens

Stefan Schmager
University of Agder
stefansc@uia.no

Charlotte Husom Grøder
Norwegian University of Science and Technology (NTNU)
charlotte.h.groder@ntnu.no

Elena Parmiggiani
Norwegian University of Science and Technology (NTNU)
parmiggi@ntnu.no

Ilias Pappas
University of Agder
ilias.pappas@uia.no

Polyxeni Vassilakopoulou
University of Agder
polyxenv@uia.no

Abstract

The adoption of Artificial Intelligence (AI) by the public sector has the potential to improve service delivery. However, the risks related to AI are significant and citizen concerns have halted several AI initiatives. In this paper we report findings from an empirical study on citizens´ attitudes towards AI use in public services in Norway. We found a generally positive attitude and identified three factors contributing to this: a) the high level of trust in government; b) the reassurance provided by having humans in the loop; c) the perceived transparency into processes, data used for AI models and models´ inner workings. We interpret these findings through the lens of social contract theory and show how the introduction of AI in public services is subject to the social contract power dynamics. Our study contributes to research by foregrounding the government-citizen relationship and has implications for public sector AI practice.

Keywords: Attitudes towards AI, trust, social contract, human-in-the-loop, transparency

1. Introduction

The introduction of Artificial Intelligence (AI) in the public sector can improve the delivery of services to citizens enabling personalization, better informed decision-making and more efficient use of resources (Pencheva et al. 2020; van Noordt and Misuraca 2022). However, AI adoption in public service delivery is still very limited with the exception of chatbots which are currently widely used as an alternative channel for information provision (Androutsopoulou et al. 2019; Aoki 2020; Mehr 2017). Extant research on the challenges of adopting AI for public service delivery has pointed to barriers related to AI-specific capabilities including capabilities for managing algorithmic performance and data governance, more general technical and managerial capabilities and regulatory hurdles (Mikalef et al. 2021; Sun and Medaglia 2019; Wirtz et al. 2019). These barriers rhyme with the ones identified for AI adoption by organizations beyond the public sector (Bérubé et al. 2021). However, AI adoption in the public sector is also challenged by the growing concerns of citizens about issues like fairness, privacy and transparency.

The public sector has to abide by the social contract which grants legitimacy to its pursuit to maximize public value for all (Rousseau 1762). This creates specific requirements for AI adoption in public services: boundary conditions for introducing AI while preserving social functions (Wilson and Van Der Velden 2022). Citizen concerns and controversies have halted several public service AI initiatives after their launch (Misuraca and Van Noordt 2020; van Veenstra et al. 2020). Recent research (Aoki 2021) has shown that concerned individuals are not ready to see decisions handled completely by AI, and public organizations have been urged to engage in democratic communications about technology with the public.

The adoption of AI in the public sector not only hinges on citizens´ attitudes about AI but also depends on their agreement to data use. Public organizations gather very large volumes of data to fulfill their missions, however, using these data to develop AI models is not straightforward. Data purpose limitation rules need to be followed and the consent of data subjects (the citizens) needs to be requested when the

URI: https://hdl.handle.net/10125/103176
978-0-9981331-6-4

boundaries of original data collection purposes are unclear and subject to interpretation. The need for obtaining such clearance is one of the reasons behind the seemingly paradoxical simultaneous overproduction and underconsumption of data by the public sector (Joseph and Johnson 2013). Providing clear information on how AI will be developed and used and ensuring citizen´s acceptance is key for AI introduction in public services.

Researchers investigating public sector AI have called for research on AI adoption examining specifically citizens´ attitudes towards AI (Saura et al. 2022; Wirtz et al. 2021; Asatiani et al. 2021). Our study responds to these calls aiming to answer the Research Question: **What are citizens' attitudes towards the use of AI in public services?** Specifically, we developed a scenario of AI use in public welfare services and conducted a study exposing participants to an interactive prototype and interviewing them to investigate their attitudes towards AI use. The study was performed in Norway during spring 2022, overall 20 participants were interviewed..

The findings of our exploratory study show a generally positive attitude towards AI in public services which the participants themselves linked to their overall trust in the government. The participants also expressed that they feel reassured by having humans involved in decisions alongside AI systems. Finally, the availability of explanations also contributes towards a positive attitude about AI, even when citizens do not fully understand the information provided. The key concerns expressed were about data collection and privacy. Our study contributes rich insights on citizens´ attitudes towards the use of AI in public services in Norway and expands extant research on public sector AI by foregrounding the government-citizen relationship. These insights also have implications for practice, as they can be used by practitioners to inform the design, development and deployment of AI systems.

2. Related Literature

2.1. AI in Public Service Delivery

The use of AI can enable public organizations to better understand and segment citizens personalizing their service offerings (Pencheva et al. 2020). Examples include tax agencies categorizing individuals and business taxpayers to tailor their services and prevent fraud, public labor organizations developing AI for profiling unemployed people to identify types of programs that are more suitable for their support and immigration authorities developing predictive AI to recommend immigration applications that merit acceptance, and spot potential red flags (Kuziemski and Misuraca 2020). Police departments have also been using AI to identify areas where they need to focus efforts to prevent crime (Höchtl et al. 2016; Waardenburg et al. 2018). However, AI applications in the context of public services have also received criticism. The Austrian labor administration created a system to categorize job seekers by their likelihood of finding a job which spurred concerns about bias and discrimination (Lopez 2021, Wimmer 2018). Overall, the main concerns expressed about public sector AI relate to training on inappropriate datasets that may perpetuate or even amplify biases (Lopez 2021) harming service impartiality (Rothstein 2013) and the opaqueness of AI inner workings (Asatiani et al. 2021; Berente et al., 2021) that limits public organizations´ability to shape, operate and monitor AI.

2.2. Explainable AI and Transparency

Explainable AI is a response to concerns about AI opaqueness and lurking biases. Explainable AI supports "meaning-making" to address questions such as "How does it work?" and "What mistakes can it make?". Explanations can help establish rapport, confidence, and understanding between AI and the user. Doran et al. (2018) assert that to achieve trustworthiness and an evaluation of the ethical and moral standards inscribed on a machine, explanations should provide insight into the rationale of the AI system and enable users to draw conclusions based on them. This is especially relevant when it comes to understanding failures and unexpected AI behavior (Ehsan & Riedl, 2019). Transparency into the inputs and inner workings of AI is considered key for its deployment in public services (Wilson and Van Der Velden 2022).

2.3. Social Contract Theory

Public service organizations execute governmental rights and obligations. They are required to provide the same services to all people regardless of social, ethnic, or religious background (Aucoin 2012). According to Jørgensen and Bozeman (2007), equal treatment of citizens, neutrality and impartiality are considered critical public values. This implies that public organizations cannot choose whether they want to offer a service or not – unlike private companies. But just like a public service organization cannot pick and choose whom they want to engage with, citizens are also bound to the public service organizations as sole providers for specific services (Junginger, 2016). These power relationships between governmental institutions and the people are part of the "social contract". According to the conceptual idea of a social contract, legitimacy to governments and institutions is assigned from the people they govern - the sovereign (Rousseau, 1762). Rousseau describes the sovereign as the collective grouping of people who by their consent enter into a civil society. As a consequence, this engenders a responsibility for the rulers to act in the interest of the ruled, but also ascribes certain rights and obligations to the latter. These specific dynamics between citizens and public service organizations, render social contract theory as a suitable lens when studying the adoption of AI by the public service. Applying this understanding of mutual duties and obligations, requires a reevaluation of privacy concerns, data-processing and transparency considerations, which are common topics in the discourse about a responsible AI implementation.

Although a social contract perspective is mostly used in political science, it is not completely foreign in the study of information systems. Smith & Hasnas (1999) examined social contract theory as a guiding theory to form "normative business ethics" in information systems for the corporate domain. Malhotra, Kim and Agarwal (2004) used social contract theory as a theoretical framework to investigate online consumer's privacy concerns. Vial (2019) suggested a turn toward normative theories of ethics such as social contract theory to research the challenges in the context of technology design and use by multiple parties.

3. Research Context & Prototype

In Norway, the government promotes the use of AI in public administration aiming to lead the way in developing human-friendly and trustworthy solutions. This study was performed in the context of a Norwegian public organization that has a central role within public administration managing different types of benefits. About five years ago, the organization established a team to explore the possibilities of data analytics and AI for delivering better, more efficient and more robust services while being committed to doing it responsibly. Among several AI initiatives the team engaged in the development of a model to predict the length of sick leaves. The purpose of this model is to become an additional resource for case handlers, helping them focus efforts where they are most needed. This links to the aim to deliver efficient services designed for "user-adapted follow-up".

This study is part of a larger research project on the responsible application of AI in the public sector which follows an Action Design Research (ADR) approach (Sein et al., 2011). Taking an ADR approach entails close collaboration with practice. In a series of iterations, we developed together with the public organization a prototype consisting of a user interface, mimicking a public service agency portal. The prototype depicted a predefined interaction sequence starting from a notification about the optional use of an AI-based prediction, different types and levels of information about the prediction and consent options.

On the first screen the prototype interface provides textual information about the legal framework for the sick leave use case. Further, it presents a value proposition, explaining the anticipated benefit for the citizen, as well as for the organization. Below the explanatory text, the interface presents links to different information elements (Figure 1).

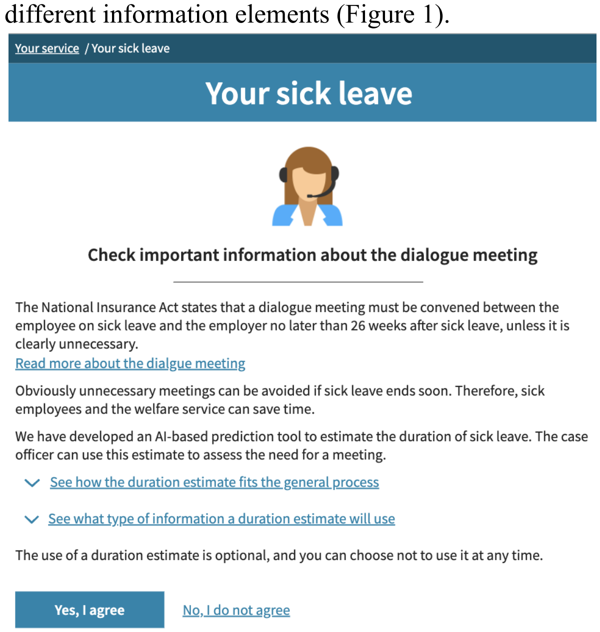

Figure 1. Prototype Interface

The first information element is a process chart aiming to provide transparency by situating the AI into the overall process and highlighting the case handler as an integral part (Figure 2).

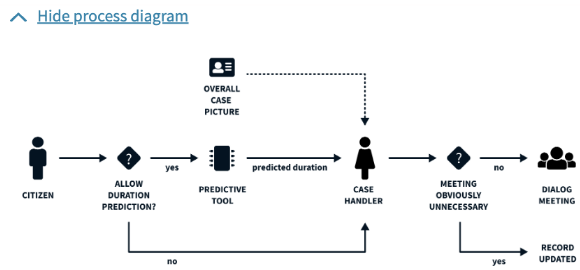

Figure 2. Process visualization

The second information element is a table providing an overview of data used by the AI system and an explanation on why this information is needed (Figure 3).

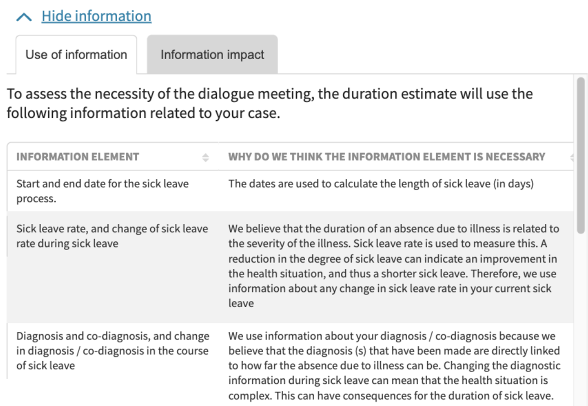

Figure 3. Data usage table

The final information element is an interactive chart, depicting relative feature importance in the form of Shapley values (Lundberg and Lee 2017), which aims to provide transparency to the model logic (Figure 4).

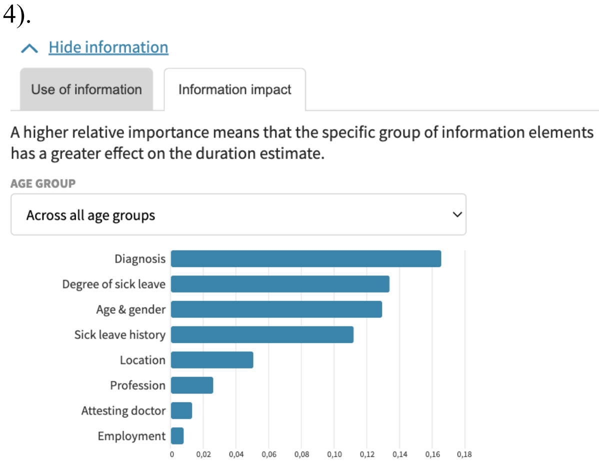

Figure 4. Feature importance charts

4. Method for Data Collection and Analysis

4.1. Recruitment & Data Collection

For this study, we recruited 20 participants between 18 and 65 years old reflecting the distribution of the general population on sick leave based on the official Norwegian statistics (Table 1) (Statistisk sentralbyrå, 2022). Further recruitment criteria were gender and education. For the gender criteria we were only able to differentiate between two genders (female and male) as these are the only defined genders in the official statistics. Moreover, we also aimed to match the statistical distribution of the educational level, including participants with high school education (Videregående), vocational school level (Fagskolenivå) and university education (Universitets- og høgskolenivå). The research sessions were conducted both, online via a video conferencing & screen-sharing application as well as in-person, depending on the availability of the participants. The study was approved by the Norwegian center for research data (NSD) and the participants gave their consent for participation.

Age Group	Number of participants
18-24	2
25-34	4
35-44	4
45-54	5
55-65	5

Table 1. Participant panel compilation

The data collection included three consecutive parts. In the initial part, we collected general data about the participants: age and gender, current occupation and highest educational level. Further, we asked the participants to provide a self-assessment for two dimensions on a scale from 1 to 5 (low - high). The first dimension was defined as *"Prior knowledge about artificial intelligence"* with an average self-reported rating of 2.20. This indicates that participants had some rudimentary knowledge about AI, but lacked deeper technical understanding. For the second dimension, *"Frequency of technology use"*, participants provided an average self-reported rating of 4.65, demonstrating exposure and general familiarity with the use of technologies like computers or smartphones. Next, we collected data on the level of trust towards the

Norwegian government and governmental organizations from the participants, as well as the use of AI technology within the public service.

In the second part, a moderated user study in the form of a task-based interaction with the prototype was conducted. Participants were presented with a short scenario and task which led them to a decision whether they would consent to the use of a new AI-supported prediction system in relation to the public service use case. While performing the task, the participants were encouraged to share their thoughts and feelings with the moderator. This "Think Aloud Protocol" method describes the concurrent verbalization of thoughts while performing a task (Ericsson & Simon, 1984). This helped us to better follow the participants line of thinking and achieve a clearer understanding of their reasonings, thoughts and concerns.

In the third and final part, we followed up with questions about the experience with the interactive prototype. After completing their interaction with the prototype, the participants were again asked to rate their level of agreement with a set of predefined statements in relation to the scenario, from 1 (strongly disagree) to 5 (strongly agree). The predefined statements were relating to the perceived competence and efficiency of an AI-infused sick leave prediction tool, the anticipation of negative consequences with the use of such a tool in public services, and the understanding of such a tool. Further, we also asked the participants about the levels of comfort and trust towards such a tool. Additionally, we asked open questions about the different explanatory elements used in the prototype and if they impacted their decision making process.

4.2. Data Analysis

To ensure a comprehensive analysis we categorized the collected data into three classes: answers to closed questions - scale ratings, spoken word feedback and interactions with the prototype. Each class of data was first analyzed individually and later synthesized with the other two classes. The goal of the analysis was to identify themes, which would help us to better understand the attitudes and concerns of participants. First, we evaluated the answers to the predefined questions given by the participants. Next, we analyzed the collected spoken word data from the Think Aloud protocol as well as the feedback the participants shared on the open ended questions about the different types of explanation. This class of data provided a wealth of useful information about participants´concerns and understandings. Lastly, we also investigated the interactions the participants performed with the prototype. Specifically, we analyzed if a specific information element was accessed, if the participant interacted with it and how long time was spent to process the information. In the final step of the analysis, we identified emerging concepts and common themes that are presented in the following section.

5. Findings

5.1. Positive Attitudes towards the use of AI

A major finding is that among the 20 participants, only one decided not to provide consent for the use of the AI tool. Interestingly, most participants expressed a rather positive stance right from the start while the prototype interaction mostly enhanced the positive attitudes towards AI.

Specifically, when the participants were asked about their agreement with the statement *"I think I would be comfortable with the use of AI for public services"*, 10/20 gave a rating of 4 or 5 and 8/20 gave a rating of 3. Only 2/20 gave a rating lower than 3. According to the given ratings we assigned the participants into three categories: skeptic (below 3); neutral (3); comfortable (above 3). After the interaction with the prototype, we followed up on this rating with the statement: *"I think I would be comfortable with the use of such a tool within public services"*, aiming to assess whether the interaction with the prototype had any effect. We found that 40% changed by increasing their comfort, approximately 40% of the participants did not change and 20 % did change by lowering their rating. Specifically, we found that after having interacted with the prototype, half of those that started neutral converted to being comfortable, one of the two "skeptics" became "neutral", and two of the "comfortables" became more comfortable. Finally, among the participants that lowered their rating one "neutral" became "skeptic" while the others remained in the "comfortable" category but reduced their expressed level of comfort.

5.2. Contributing Factors to Positive Attitudes towards AI

5.2.1 Trust in Government. Exploring the generally positive attitude towards AI in public services we found that the participants linked it to their overall trust in the government. Several participants provided revealing articulations about their trust to government and how this trust affects their choice to consent to the use of the AI system: *“I don’t have time to read all this, but I trust the government”; “If that would be like a private company, e.g. if it’s [name of a telecom company], I trust them somewhat, but these companies sell the information given to other companies for marketing and whatever.”* These statements indicate a deep belief in the integrity of the Norwegian government.

However, some of the participants also expressed reflective comments on how self-aware they are about their level of trust to the Government: *“[...] I'm a typical Norwegian, super naïve to the government. I really think they are trying to do the best, even if they don’t.”; “Maybe I'm naive, but I trust the government.”* and *“I trust the government and I think they are trying to do the best for the people in Norway.”* This sentiment shows a clear self-awareness about the high level of trust in the government which manifests a feeling of comfort but also potentially a concern about ensuring that their trust is not betrayed.

5.2.2 Being reassured by having humans in the loop. The next identified theme in our investigation about the generally positive attitude towards AI concerns the role of human actors during the process. The participants expressed that they feel reassured by having humans involved in decisions alongside AI systems. Within our public service scenario, the calculated sick leave duration prediction is used as an additional information source for the case worker. It is not an Automated Decision Making system. The general concept of having a human involved during the process, has been brought up by participants in different variations. One of the participants said: *“I will be comfortable at least if it's not only the tool that's a part of the decision, but with people”*. Another participant stated: *“It would be much more easy for me if the element of the “case handler” was more visible”* One participant explained that a fully automated system would be a concern: *"[...]decisions are made with too little discretion because they fully trust the system"*. The common sentiment from these statements can be understood as an implicit expectation that a machine should not be left alone to make a decision, but a human actor would still retain the decision power. However two participants did express their wish to get humans out of the loop to ensure impartiality. One of them said: *“I think some people are very pushy and begging, and maybe they get more. And some people don't ask for much, so they can have real problems and in a way to me it's more fair if it's based on this [AI system] not if I'm yelling or crying”*. This hints towards a perception that having a human within the process may also be a potential weakness that could lead to a less fair or equal treatment of citizens.

Some participants reasoned about having a human in the loop relating this to their need to ensure “contestability”. Hence, contestability was identified as a separate theme in our empirical material described as the capability to argue about changing a decision when perceived to be wrong. Participants expressed the concern that if a decision would be made without a human involved in the process, it would be difficult for them to find a person to raise such a dispute. This was expressed explicitly by one participant as *“Where should I complain to?”* Another participant explained: *“AI's rating may be incorrect. With a person, you can explain what is wrong, but you can not to AI”*.

5.2.3 Transparency. The final theme identified in our analysis relates to the transparency provided by the prototype. Although the overall impression about the explanations provided was positive, and the prototype interaction mostly enhanced the positive attitudes towards AI, participants also provided several comments which indicate that AI remained relatively opaque for many. Multiple participants commented on difficulties understanding the text. This was attributed to the use of concepts that require some basic understanding of statistics: *“Maybe like this sentence in here, it's probably a bit difficult to understand for all people that are not working with statistics”*. Also, participants remarked that some of the wording was difficult to understand due to its legislative content: *“I don’t think normal people would understand this; it’s like a lawbook”* and *“Heavily written, very classic bureaucratic language - "duration estimate" I do not think many people understand that much of”*. But also specific words were either perceived as hard to understand: *“I don't understand the word information effect”*.

The process chart received mixed feedback. For some participants, the chart helped to understand how the system would be used within the overall context and which role it would play in relation to the general process: *“I really love this”; “I love to see how the process is with and without the tool”*. Others found the process chart difficult to understand: *“Hard to figure out the chart”* or *“I don’t understand anything about this”*

Similarly, although some participants found the data table useful: *"Yes, I think it is useful that it says what the purpose of collecting information is, what kind of information is collected",* other participants found the data table excessive, mentioning: *“I would* *just close it right away 'cause I would think I don't have the time to read this or understand it so I wouldn't read it."* or *"Too much to read"*. Interestingly, some of the participants expressed surprise when they saw the data table, as they did not expect that the government would have all this data. Further, some participants also wondered who else might have access to the data: *"Do they have this information?"; "Will my employer see this information?"* and *"I feel an employer can use it against me and in further hiring processes. As an employee you already have a weaker position."*

These findings may indicate that the sole existence and availability of information and explanations contribute towards a positive attitude about AI, although citizens do not fully understand the information provided.

6. Discussion

The findings of our study show a generally positive attitude towards the use of AI in public services. A deeper analysis of the empirical material led to the identification of three factors contributing to this positive attitude. These factors are: a) the high level of trust in government; b) the reassurance provided by having humans in the loop; c) the perceived transparency into processes, data used for AI models and the inner workings of the models. In the following paragraphs, we discuss these findings through the social contract lens applied in this research study.

Believing that the government, and its institutions act in the interest of the citizens is an integral part of social contract theory and can be seen as a fundamental requirement for a society to work. It enables the rulers to fulfill governmental duties and to retain legitimation. In our context of being asked to consent to the use of AI for public services, the trust to government expressed by the participants describes the expectation of a responsible use of AI and processing of data. In particular, it relates to the expectation of transparency on which data will be used, what is the rationale of the data usage and for which benefits and purposes, even if the provided transparency and explanations are not fully understood. Citizens trust that their consent will be used by the authorized organization for its intended purposes and to their benefit which can be seen as an implied social contract (Perreault 2015). This social contract is considered breached, if citizens are unaware of unauthorized data collection or data processing or data transfer to other parties without their explicit consent (Xu et al. 2006).

This also touches on the expectation that new AI enabled processes will include safeguards, e.g. by having humans in the loop during decision making processes, providing possibilities to reach out to a human actor for help or to contest a specific outcome. These findings point towards an implicit expectation from the individual to be able to exercise their right as the sovereign to question governmental decisions. The underlying assumption is a working relationship under the dynamics of the social contract, which makes participants content to allow the use of AI.

In the context of a responsible AI implementation, this condition introduces multilateral implications. On the one hand, it creates the opportunity for governmental organizations for a sensible and beneficial introduction of AI systems. Leveraging such technologies can benefit the organization by enabling them to enhance services, increase efficiency and improve processes. In turn, the consent of the people to allow governmental organizations use AI, can contribute to benefits for the individuals and society at large. But these statements also highlight a peril that a high level of trust can entail. As governments and governmental organizations obtain their right to rule and govern by the consent of the public, this also requires the latter to critically scrutinize governmental decisions. Especially in the context of AI in the public sector, the risks are significant. At the same time, it is a contractual responsibility of the government and its organizations to not only avoid, but also to prevent any misuse and to safeguard citizens from exploiting the granted trust. The realization of a breach of trust can lead to halting public service AI initiatives after their launch (Misuraca and Van Noordt 2020; van Veenstra et al. 2020).

Interpreting the theme of "human-in-the-loop", from a social contract perspective alludes to the perception that a human actor would be considered a more appropriate contractual partner than a machine. The level of trust on the overall benevolence of the public service organization seems to be extended towards the public service employee, however less to the AI system. Several of the participants understand the role of the human as a safety measure to prevent unfair treatment. Having a human retaining the ultimate power of decision making, being available for help or an authority for objection describes a role in the contractual relationship to which the people attribute certain responsibilities to act in their interest as part of the general will. However, this expectation creates a potential for tensions if the general will is not aligned with the interests of the individual. Interestingly, those participants mentioning an improved objectivity of the AI system seem to be concerned about the same point, as they are expecting a fairer and more equal treatment being the result of an impartial AI system.

AI adoption in the public sector entails abiding by the social contract which grants legitimacy to its pursuit to maximize public value for all (Rousseau 1762). This creates requirements for the boundary conditions for introducing AI while preserving social functions (Wilson and Van Der Velden 2022). Our findings are consistent with recent research (Aoki 2021) that has shown that concerned individuals are not ready to see decisions handled completely by AI, and public organizations have been urged to engage in communications about technology and provide the assurance of having humans involved in decisions alongside AI systems.

6.1. Contribution to Research

Our findings expand extant research on AI adoption in the public sector (Mikalef et al. 2021; Sun and Medaglia 2019; Wirtz et al. 2019) by foregrounding the government-citizen relationship that has received limited attention in this body of literature. The overall positive attitudes of citizens in this study align with prior research which shows that although AI can be a source of public anxiety, informing citizens about the characteristics of the AI system and of the humans´ involvement in decisions has a significant positive influence on the public's attitude towards AI which is important for its adoption (Aoki 2021). However, by being able to probe participants to share their reflections, our research goes beyond prior survey-based research (Aoki 2021) providing insights on their reasoning behind their attitudes. The information provided to citizens fortifies their trust to the government and having a human in the loop reassures people that it is possible for them to explain their particular circumstances and even contest algorithmic suggestions if needed.

6.2. Contribution to Practice

This study provides insights into how important it is for public organizations to ensure that public´s goodwill is not eroded. This can happen if public-sector projects go beyond boundary conditions for introducing AI while preserving social functions (Wilson and Van Der Velden 2022). The erosion of people´s goodwill can limit the ability of public organizations to deliver their services effectively in the future. By having identified contributing factors for a positive attitude towards AI we provide practitioners with hands-on pointers when considering a responsible and sustainable design and development of AI systems in the public service context.

6.3. Limitations and Further Research

This study is exploratory in nature. It responds to calls for research on citizens´ attitudes towards AI (Saura et al. 2022; Wirtz et al. 2021; Asatiani et al. 2021) drawing from rich empirical data collected using a combination of closed and open questions and observations of participants´ interactions with a prototype. The 20 participants of the study have been carefully selected to represent the population relevant to the AI application studied. However, to unpack the attributes of citizens´ attitudes and explore their interplay, focused follow-up studies with larger numbers of participants are required. Moreover, starting from the findings of this study, further studies can extend longitudinally over time and across countries. The stances of citizens and their acceptance of AI depend on their cultural identity and especially their overall trust to government which is particularly high in Norway (OECD 2022). Therefore, it is important to not only perform further focused research with increased samples, but also, to conduct research in different sociocultural environments. As identified in prior research, trust is systemic in nature, invested in the larger system of public and private actors that are associated with AI (Steedman et al. 2020; Wilson and Van Der Velden 2022). This is an exciting research opportunity for collaboration between international research partners interested in exploring and developing a human-centered AI framework for public services. The key role of trust for AI adoption signifies the need for active research on the mechanisms for trust building not simply asking for trustworthy AI but actually operationalizing what is trustworthy for citizens. The need for further research in this direction rhymes with recent research that pointed to the perils of trust commodification marked by a decreasing understanding of trust as the expected reliability, and an increasing emphasis on instrumental framings of trust as a resource obscuring the mechanisms through which trust in AI systems might be built, making it less likely (Krüger and Wilson 2022).

This research explores the perspectives of citizens. A key finding is that citizens feel reassured when decisions are supported by AI but not fully automated (i.e. when public servants are included in the loop). This is a finding relevant for the design of AI-enabled public services, however, a more complete picture can be developed by exploring both citizens´ and public servants´ stances regarding digital discretion in AI-supported public services. Taking a social contract perspective can expand prior research on digital discretion (Busch 2019) by shifting attention from the logics of public servants to the dynamics between citizens and public servants.

7. Conclusion

Our study provides insights on citizens´ attitudes towards the use of AI in public services in Norway. We found a generally positive attitude towards AI and identified three factors contributing to this positive attitude. These factors are: a) the high level of trust in government; b) the reassurance provided by having humans in the loop; c) the perceived transparency into processes, data used for AI models and the model inner workings. By interpreting the findings through the lens of social contract theory, we can provide an explanation of these factors´significance. Inherent trust into a government and its institutions lays the foundations to assume best intentions for the greater good. Human involvement and availability during processes facilitates the power dynamics between the rulers and the ruled, enabling the exercising of rights and obligations. Transparency into processes, data collection and data use are considered important on a cursory level, as the availability of explanations is deemed more relevant than a thorough understanding. By providing new insights into the contributing factors for a positive attitude towards AI, we hope to advance the discourse on the responsible adoption of AI in the public sector.

8. References

Androutsopoulou, A., Karacapilidis, N., Loukis, E., and Charalabidis, Y. (2019). "Transforming the Communication between Citizens and Government through Ai-Guided Chatbots," Government Information Quarterly, 36 (2), 358-367.

Aoki, N. (2020). "An Experimental Study of Public Trust in Ai Chatbots in the Public Sector," Government Information Quarterly, 37(4), 101490.

Aoki, N. (2021). "The Importance of the Assurance That “Humans Are Still in the Decision Loop” for Public Trust in Artificial Intelligence: Evidence from an Online Experiment," Computers in Human Behavior (114), 106572.

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). "Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems." Journal of the Association for Information Systems, 22(2), 325-352.

Aucoin, P. (2012). "New political governance in Westminster systems: Impartial public administration and management performance at risk". Governance, 25(2), 177-199.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). "Managing artificial intelligence". MIS Quarterly, 45(3), 1433-1450.

Bérubé, M., Giannelia, T., and Vial, G. (2021). "Barriers to the Implementation of AI in Organizations: Findings from a Delphi Study," Proceedings of the 54th Hawaii International Conference on System Sciences

Busch, P. A. (2019). Digital discretion acceptance and impact in street-level bureaucracy. PhD Doctoral Dissertation. University of Agder, Norway.

Doran, D., Schulz, S., & Besold, T. R. (2018). What does explainable AI really mean? A new conceptualization of perspectives. . CEUR Workshop Proceedings,

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. the MIT Press.

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: towards a reflective sociotechnical approach. In the International Conference on Human-Computer Interaction.

Höchtl, J., Parycek, P., and Schöllhammer, R. (2016). "Big Data in the Policy Cycle: Policy Decision Making in the Digital Era," Journal of Organizational Computing and Electronic Commerce 26 (1-2), 147-169.

Jørgensen, T. B., & Bozeman, B. (2007). "Public values: An inventory". Administration & society, 39(3), 354-381.

Joseph, R.C., and Johnson, N.A. (2013). "Big Data and Transformational Government," IT Professional 15(6), 43-48.

Junginger, S. (2016). Transforming Public Services by Design: Re-orienting policies, organizations and services around people. Routledge.

Kuziemski, M., and Misuraca, G. (2020). "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings," Telecommunications policy, 44(6), 101976.

Krüger, S., & Wilson, C. (2022). "The problem with trust: on the discursive commodification of trust in AI." AI & SOCIETY, 1-9.

Lopez, P. (2021). "Bias does not equal bias: A socio-technical typology of bias in data-based algorithmic systems." Internet Policy Review, 10(4), 1-29.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017).

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). "Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model." Information systems research, 15(4), 336-355.

Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government. Ash Center for Democratic Governance and Innovation. Harvard Kennedy School. Cambridge, MA.

Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S.O., Torvatn, H.Y., Gupta, M., and Niehaves, B. (2021). "Enabling AI Capabilities in Government Agencies: A Study of Determinants for European Municipalities," Government Information Quarterly, 101596.

Misuraca, G., and Van Noordt, C. (2020). "AI Watch-Artificial Intelligence in Public Services: Overview of the Use and Impact of Ai in Public Services in the EU," JRC Working Papers:JRC120399.

OECD (2022), Drivers of Trust in Public Institutions in Norway, OECD Publishing, Paris, https://doi.org/10.1787/81b01318-en.

Perreault, Lucille, (2015). "Big Data and Privacy: Control and Awareness Aspects" CONF-IRM 2015 Proceedings.

Pencheva, I., Esteve, M., and Mikhaylov, S.J. (2020). "Big Data and I–a Transformational Shift for Government: So, What Next for Research?," Public Policy and Administration, 35(1), 24-44.

Rothstein, B. (2013). "Corruption and social trust: Why the fish rots from the head down." Social Research, 80(4), 1009-1032.

Rousseau, J. J. (1762). The social contract (1998 edition). Hertfordshire: Wordsworth Classics of World Literature.

Saura, J.R., Ribeiro-Soriano, D., and Palacios-Marqués, D. (2022). "Assessing Behavioral Data Science Privacy Issues in Government Artificial Intelligence Deployment," Government Information Quarterly, 101679.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). "Action design research." MIS Quarterly, 35(1), 37-56.

Smith, H. J., & Hasnas, J. (1999). "Ethics and information systems: The corporate domain." MIS Quarterly, 109-127.

Statistisk sentralbyrå (2022). Sickness absence. Retrieved 2022 from:https://www.ssb.no/en/arbeid-og-lonn/arbeidsmiljo-sykefravaer-og-arbeidskonflikter/statistikk/sykefravaer.

Steedman, R., Kennedy, H., and Jones, R. (2020). "Complex Ecologies of Trust in Data Practices and Data-Driven Systems," Information, Communication & Society 23(6), 817-832.

Sun, T.Q., and Medaglia, R. (2019). "Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare," Government Information Quarterly, 36(2), 368-383.

van Noordt, C., and Misuraca, G. (2022). "Artificial Intelligence for the Public Sector: Results of Landscaping the Use of Ai in Government across the European Union," Government Information Quarterly), 101714.

van Veenstra, A.F., Grommé, F., and Djafari, S. (2020). "The Use of Public Sector Data Analytics in the Netherlands," Transforming Government: People, Process and Policy.

Vial, G. (2021). "Understanding digital transformation: A review and a research agenda." Managing Digital Transformation, 13-66.

Waardenburg, L., Sergeeva, A., and Huysman, M. (2018). "Hotspots and Blind Spots," Working Conference on Information Systems and Organizations.

Wilson, C., and Van Der Velden, M. (2022). "Sustainable Ai: An Integrated Model to Guide Public Sector Decision-Making," Technology in Society (68), 101926.

Wimmer, B. 2018. Der Ams-Algorithmus ist ein „Paradebeispiel für Diskriminierung". Retrieved February 22, 2022 from https://futurezone.at/netzpolitik/der-ams-algorithmus-ist-ein- paradebeispiel-fuer-diskriminierung/400147421

Wirtz, B.W., Langer, P.F., and Fenner, C. (2021). "Artificial Intelligence in the Public Sector-a Research Agenda," International Journal of Public Administration 44(13), 1103-1128.

Wirtz, B.W., Weyerer, J.C., and Geyer, C. (2019). "Artificial Intelligence and the Public Sector—Applications and Challenges," International Journal of Public Administration, 42(7), 596-615.

Xu, H., Teo, H. H., & Tan, B. (2006). Information privacy in the digital era: an exploratory research framework. AMCIS 2006 Proceedings..