Data & Policy (2024), 6: e19
doi:10.1017/dap.2024.13

RESEARCH ARTICLE

Exploring citizens' stances on AI in public services: A social contract perspective

Stefan Schmager[1], Charlotte Husom Grøder[2], Elena Parmiggiani[2,3], Ilias Pappas[1,2] and Polyxeni Vassilakopoulou[1]

[1]Department of Information Systems, University of Agder, 4630 Kristiansand, Norway
[2]Department of Computer Science, Norwegian University of Science and Technology, 7491 Trondheim, Norway
[3]Sintef Nord AS, 9008 Tromsø, Norway
Corresponding author: Stefan Schmager; Email: stefan.schmager@uia.no

Received: 22 June 2023; **Revised:** 06 November 2023; **Accepted:** 20 February 2024

Keywords: agreement modeling; human-in-the-loop; social contract; stances toward AI; trust

Abbreviations: AI, Artificial intelligence; ADR, Action Design Research; NSD, Norwegian center for research data; SCT, Social Contract Theory

Abstract

This paper explores citizens' stances toward the use of artificial intelligence (AI) in public services in Norway. Utilizing a social contract perspective, the study analyzes the government–citizen relationship at macro, meso, and micro levels. A prototype of an AI-enabled public welfare service was designed and presented to 20 participants who were interviewed to investigate their stances on the described AI use. We found a generally positive attitude and identified three factors contributing to this: (a) the high level of trust in government (macro level); (b) the balanced value proposition between individual and collective needs (meso level); and (c) the reassurance provided by having humans in the loop and providing transparency into processes, data, and model's logic (microlevel). The findings provide valuable insights into citizens' stances for socially responsible AI in public services. These insights can inform policy and guide the design and implementation of AI systems in the public sector by foregrounding the government–citizen relationship.

Policy Significance Statement

As artificial intelligence (AI) technology evolves, it holds great potential for enhancing public service delivery. But citizen concerns underscore the need for responsibility and diligence in the public sector. To understand citizen perspectives on AI use in Norwegian public services, we examined a public welfare case and analyzed the results through a social contract theory lens. This perspective considers government–citizen relationships at multiple levels. We found three key factors driving positive citizen views of AI: (1) high government trust (macro level); (2) balancing individual and collective needs (meso level); and (3) confidence in human oversight and transparent processes, data, and model logic (microlevel). Our insights can guide policymakers in responsibly integrating AI for the benefit of both individuals and society.

1. Introduction

There is strong potential for artificial intelligence (AI) to revolutionize how governments deliver services to citizens. AI technologies can enable the delivery of personalized services, better inform decision-making, and contribute to more efficient use of resources (Pencheva et al., 2020; van Noordt and Misuraca, 2022). However, AI adoption in public service delivery has so far been relatively slow and narrow in scope. Chatbots for information provision represent the primary—albeit limited—application of AI in public services (Mehr et al., 2017; Androutsopoulou et al., 2019; Aoki, 2020). More advanced and high-impact uses of AI—from predictive analytics to AI-assisted decision-making—have seen little real-world implementation in the context of public services.

The purpose of public organizations is to mediate the relationships between government and citizens and make positive contributions to society, by providing their services to citizens as well as instruments for implementing public policies (Junginger, 2016). The public sector has to abide by the social contract that grants legitimacy to its pursuit to maximize public value for all (Rousseau, 1964). This creates specific requirements and boundary conditions for adopting AI in public services while at same time, preserving social functions (Wilson and Van Der Velden, 2022). Citizens expect governments to demonstrate transparency, accountability, and safeguards that address issues of fairness, privacy, and bias before endorsing the use of AI for public service delivery. This is evidenced by multiple cases of public service AI initiatives that were halted after their launch due to citizen concerns and controversies (Misuraca and Van Noordt, 2020; van Veenstra et al., 2021). Furthermore, prior research (Aoki, 2021) has shown that concerned individuals are not ready to see decisions about them handled completely by AI, and public organizations have been urged to engage in democratic communications about technology with the public. Realizing the potential of AI will require governments to ensure acceptance by citizens addressing their concerns before AI systems are launched.

The adoption of AI in public services also depends on citizens' agreement for the reuse of their data for training AI models. Public organizations gather large volumes of data to fulfill their missions; however, using these data to develop AI models is not straightforward. Especially for personal data, in Europe, data purpose limitation rules need to be followed and the consent of data subjects (i.e., the citizens) needs to be requested when the boundaries of original data collection purposes are unclear and subject to interpretation (EU, 2016). Problems can arise when data are collected for one purpose and later used for another (Verhulst, 2021). 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).

We propose the adoption of a social contract perspective (Rousseau, 1964) for exploring AI in public services. A social contract perspective can provide a better understanding of the dynamics in government–citizen relationships supporting the identification of important requisites for AI introduction and the anticipation of tensions that may arise. Specifically, we suggest three different social contract-based lenses on government–citizen relationships at macro, meso, and micro levels. The lenses are useful for analyzing and critiquing existing AI initiatives, and also for proactively designing more legitimate and acceptable visions of AI for public services that uphold—rather than erode—fundamental elements of the social contract. In this paper, we apply the three lenses on a prototype for an AI-enabled public service to citizens, which was used as a probe for eliciting the stances of citizens toward the use of AI in public services. To collect empirical data, we developed a scenario of AI use in public welfare services and conducted a qualitative study exposing participants to the interactive prototype and interviewing them to investigate their stances toward the described AI use. The study was performed in Norway during 2022, with 20 participants being interviewed.

We found that citizens are generally positive and identified three factors contributing to this: (a) the high level of trust in government (macro level); (b) the balanced value proposition between individual and collective needs (meso level); and (c) the reassurance provided by having humans in the loop and providing transparency into processes, data and model's logic (microlevel). The findings provide valuable insights into citizens' stances on socially responsible AI in public services. Our study contributes rich insights into citizens' stances toward 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 have implications for practice, as they can be used to inform policy and also, the design and deployment of AI systems.

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

2. AI in public service delivery

As technology advances at a rapid pace, there are calls for urgent conversations about the responsible use of these technologies (Future of Life Institute, 2023; Center for Humane Technology, 2023). Private technology companies like Microsoft, Google, and others seem to be in an arms race to develop and deploy one AI breakthrough after another. There have been various undertakings to formulate normative guidelines for the responsible design and development of AI with the goal of human benefit (Xu, 2019; Shneiderman, 2020). Concepts like "human-in-the-loop" (HITL) are getting established as mechanisms for AI systems that better serve human needs (Zanzotto, 2019). An HITL mechanism describes systems in which human involvement is integrated with automated processes to improve overall performance, reliability, and decision-making (Fu et al., 2021; Macher et al., 2021; Herrmann, 2022). Furthermore, several technology companies have defined their own sets of responsible AI guidelines with a human-centered foundation (Wright et al., 2020). However, most of these guidelines are created with a profit-oriented premise, inherited by their commercial origin, conceiving the interaction between the involved actors as a client-vendor relationship.

There are significant opportunities provided by AI systems in public services. Governments can profit from AI technologies as they have potential access to extensive amounts of data that can be harnessed for AI system development. There are multiple examples of the potential use of AI systems in the public service context, for instance, to perform comprehensive and accurate predictions, detect fraud, or use natural language processing to process information (de Sousa et al., 2019; Misuraca et al., 2020; Misuraca and Van Noordt, 2020). The use of AI can enable public organizations to better understand and serve citizens by personalizing their offerings (Perreault, 2015; van Noordt and Misuraca, 2022). Examples of real-world applications 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).

The responsible development and use of AI in public services entail harnessing its power while minimizing risks for individuals and society (Vassilakopoulou et al., 2022). 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 (Sun and Medaglia, 2019; Wirtz et al., 2019; Mikalef et al., 2022). These barriers rhyme with the ones identified for AI adoption by organizations beyond the public sector (Bérubé et al., 2021). Schmager (2022) found that existing approaches cannot be directly translated to the public sector context and need to be adapted and individually scrutinized. Furthermore, Benedikt et al. (2020) examined the applicability and intricacies of having HITL within the public sector context and determined that in situations where a balance between efficiency and data quality need to be balanced, human interventions are needed.

Researchers investigating public sector AI have called for research that examines specifically citizens' stances toward AI (Wirtz et al., 2019; Asatiani et al., 2021; Saura et al., 2022). To do so, it is important to consider the special relationship between government and citizens. AI systems for public services should be designed and implemented in ways that do not infringe on core values and citizen rights (Dignum, 2019). To elicit and understand such values, a notable guiding principle is to ensure public engagement as well as discourse with society. Citizens expect government services to fulfill public missions and implement policies in the interest of society as a whole upholding the legitimacy granted by the social contract between governments and citizens. A social contract (Rousseau, 1964) perspective can provide insights that are specific to the introduction of AI in the public sector.

2.1. The challenge of socially responsible AI in the context of government—Citizen relationships

The special relationship between the public—the ruled—and public governance—their rulers—makes the use of AI technologies in the public sector particularly sensitive. Public service organizations, as the

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

name suggests, are under a duty to serve the public. They obtain their legitimization by the consent of the public, which implies their interests should be driven by the benefit of society and the collective good—the general will (Rousseau, 1964).

Citizen concerns and controversies have halted several public service AI initiatives after their launch (Misuraca and Van Noordt, 2020; van Veenstra et al., 2021). There are multiple examples of cases where the use of AI in the context of public services has 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 (Wimmer, 2018; Allhutter et al., 2020; Lopez, 2021). The Dutch government developed an AI system that links multiple data sources and provides indications of possible fraud by welfare beneficiaries. In response to that, civil society organizations and individuals convened and filed a freedom of information request asking questions about its workings and use (Wieringa, 2023). The Dutch courts found that the system violates citizens' rights by being untransparent and challenging privacy (Bekker, 2021). In their examination of possible adverse consequences of algorithmic decision-making (ADM) within the public sector, Rinta-Kahila et al. (2022) investigated an ADM system implemented by the Australian government to automatically calculate and collect welfare overpayment debts from citizens. They found that the system inadvertently led to significant distress among both citizens and employees of the public service organization. The destructive effects ultimately harmed the government too, both financially and in terms of its reputation and citizen trust. Overall, key 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 blackboxing AI's inner workings (Asatiani et al., 2021). This urgently raises the issue on how we can ensure socially responsible and legitimate use of AI in public services by taking into account the special relationship between the citizens and their government.

In the exploration of theoretical frameworks for investigating socially responsible AI in public services, multiple alternative theoretical lenses were considered. Specifically, three promising theories were examined, including stakeholder theory, institutional theory, and structuration theory. Stakeholder theory can be viewed as inherently managerial; a framework for managerial and organizational behavior (Freeman, 1984). It has been applied in the field of public service research but has received varying appraisals. Although Flak and Rose (2005) found no conceptual mismatch between stakeholder theory and the government's objective of providing policy and services for citizens and organizations, Donaldson and Preston (1995) describe the theory as merely one of the private sector, governed by fundamentally different principles and implications than any public sector organization. Its key limitation, compared to social contract theory, is its limited capacity to comprehend the specific roles and reciprocal obligations in a governmental context beyond executive responsibilities. We also examined institutional theory and structuration theory that explore the creation and enactment of structures, their relationship with context, actions, and actors (Scott, 2004; Giddens, 2014). Institutional theory suggests organizations strive for legitimacy to ensure long-term survival, which is a concept aligned with social contract theory (Meyer and Rowan, 1977). Meanwhile, structuration theory recognizes how actors operate within the constraints of social structures, offering insights into power dynamics, legitimacy, and organizational responsibilities (Cohen, 1989). However, as this study is not focused on structure dynamics but rather on socially responsible AI in the context of citizen—government relationships, social contract theory was deemed to be the more suitable theoretical framework.

2.2. Need to reposition AI in public services: A social contract theory perspective

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

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

sole providers for specific services (Junginger, 2016). These power relationships between governmental institutions and the people are part of the "social contract."

In this research, we take a Social Contract Theory perspective for socially responsible AI in public services to explore the relationships and power dynamics between the involved actors. The term Social Contract Theory encompasses a large body of theories and should be understood as a collection of conceptual streams rather than a single specific theory. In general, a social contract describes a hypothetical agreement in which the members of a society endorse and comply with the fundamental rules, institutions, and principles of that society. It is an agreement made between citizens and the state, through which they accept the authority of the state in return for benefits that only a sovereign power can provide, delineating a structure of power relationships between governments, institutions, and the people (Heywood, 2015). Power can be understood as the capacity to make formal decisions that are binding on others (Heywood, 2015). According to D'agostino (1996), a social contract is concerned with determining whether a ruling power is legitimate and therefore worthy of loyalty. In 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 (Figure 1).

The social contract formalizes the effort to create consensus on shared values and ideals and ensure ethical practice (Jos, 2006). The ultimate aim of a social contract approach is to demonstrate that social, moral, or political rules can be rationally justified. According to Rawls (2001), the social contract is a model of rational justification that transforms the problem of justification into a problem of deliberation. It consists of general parameters and dynamics, set to represent reasons for endorsing and complying with social rules, principles, and norms. Yet justification does not rely on some exogenous reason or truth but is rather generated endogenously by a form of agreement. 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.

According to the idea of a social contract by Rousseau (1964), legitimacy to governments and institutions is assigned from the people they govern—the sovereign. Rousseau describes the sovereign as the collective grouping of people who by their consent enter into a civil society. David Gauthier (1986) argues that any system of moral constraints must be justified to those to whom it is meant to apply. Applying this understanding of mutual duties and obligations requires a reevaluation of privacy concerns,

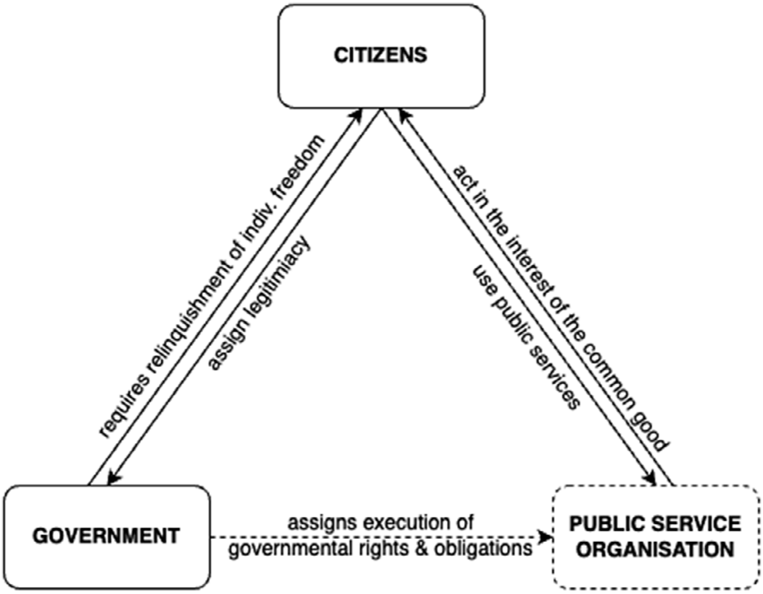

CITIZENS
requires relinquishment of indiv. freedom
assign legitimacy
act in the interest of the common good
use public services
GOVERNMENT
assigns execution of
governmental rights & obligations
PUBLIC SERVICE
ORGANISATION

Figure 1. *Relationships within the social contract.*

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

data processing and transparency considerations, which are common topics in the discourse about responsible AI implementation.

Citizens expect that their data 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 points toward an implicit expectation from the individual to be able to exercise their right as the sovereign to question governmental decisions. Furthermore, this also examines how a social contract, which is not a formalized contract but rather a conceptual one, can be broken. For example, Xu et al. (2006) found that a social contract in the case of technology use in the public sector is considered breached if citizens are unaware of unauthorized data collection or data processing, or data transfer to other parties without their explicit consent, which illustrates a form of power abuse. This level of examination can be understood as the macro-lens of social contract theory.

At the next level, which we label as the meso-lens, one needs to consider the interaction between the individual and society. This relationship has been touched by almost all political debates (Heywood, 2015). Applied within the research of socially responsible AI, it can help to better investigate the balancing act between individual rights and benefits, and the common good. If we stick with the example of personal data, which reasons do individuals have to share personal data, which provides benefit for everyone within a society? Why are they prioritizing either societal cooperation or personal gain? These reasons individuals have for agreeing to some rules or principles need to be understood as their own reasons, not necessarily as "good reasons" from the impartiality perspective. Individuals may care about what they perceive to be impartially good or some other non-individualistic notion, but what they care about will differ from one another. Rawls (1996) highlights that a society cannot reasonably be expected to have similar conceptions of the good, and heterogeneity needs to be considered. And the same is true for the perspective of the government organization. Does the organization respect individual freedom, or does it exercise its power for the benefit of the collective? If a social contract is to endorse interrelated normative desiderata (e.g., liberty, equality, and welfare, which are all guiding principles for a working society), a deliberative process that draws on a diversity of perspectives will outperform one based on a strict normalization of perspectives. A human-centered approach that acknowledges and incorporates the tensions between the collective good and the individual good forms a suitable conceptual framework.

The final tier of examination within the framework of social contract theory is found in the micro-lens. This level explores the intricate processes and mechanisms that underpin the attainment of consensus among the various involved stakeholders. It places particular emphasis on the dynamics and interactions between citizens and public organizations, scrutinizing the immediate communication and exchanges among these different actors to construct an agreement. Among the most widely discussed mechanisms are consent (Rousseau, 1964; Hobbes, 2016; Locke, 2016), bargaining (Nash, 1950; Harsanyi, 1976), aggregation (Harsanyi, 1976), and equilibrium (Buchanan, 1975; Gauthier, 1986). These mechanisms delineate various forms of normative authority for self-binding, predicated on the notion that individuals possess fundamental normative powers over themselves. If parties can indeed bind themselves by wielding this normative power, then the outcome of the social contract results in an obligation (as asserted by Hobbes, 2016; Hume, 1963). It is worth noting that some of these agreement mechanisms have faced scrutiny from contemporary social contract theorists, with consent, in particular, standing out as a key example (Weale, 2020).

3. Research context and methodology

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 in 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

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

handlers, helping them focus efforts where they are most needed. This links to the aim to deliver the efficient services designed for “user-adapted follow-up.”

3.1. Prototype

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 an interactive prototype consisting of a user interface, mimicking a public service agency portal (Table 1). The prototype depicts 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.

The development of the prototype was driven by the social contract lenses framework, which informed the design decisions on multiple levels. On the macro-level, the prototype needed to take the existing power structures into account and reflect on the rights and duties of each actor in this process. The existing rights and duties of both parties—the governmental organization as well as the citizens—needed to be presented in a sufficient manner. One element in this specific case was the transparency and acknowledgment of which personal data about the citizen would be accessible to the organization per the nature of the organizational power, in case of an agreement to use. Simultaneously, it also needed to be clear to the citizen that the right to object to the use of this personal data exists without any negative consequences or need for justification. On the meso-level, the participation and contribution of an individual to the collective good by agreeing to the use of personal data should be outlined. In particular, this needs to address deliberation of social morality, rather than the moral obligation of the citizen but also requires trust in the honesty and integrity in the public service organization. Doran et al. (2017) 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. In this specific use case, the purpose, and anticipated benefits of such a system for the general society, as well as in particular to the individual need to be presented. For the individual citizen, the benefit would result in the potential avoidance of unnecessary appointments. For the organization, the use of personal data, and in consequence, the use of the algorithmic decision-support system would result in improved processes, and by that, more efficient resource utilization. Finally, on the microlevel, the prototype had to support the immediate user–system interaction to consent or dissent to the request. This means that the interaction and communication sequence needed to enable the users (the citizens) to understand their role in the process, grasp the governmental request and the consequences of either agreement or disagreement, and act accordingly.

3.2. Recruitment

For this study, we recruited 20 participants aged between 18 and 65 years, reflecting the distribution of the general population on sick leave based on the official Norwegian statistics office (Statisisk Sentralbyrå, 2022) (Table 2). The number of participants was defined considering the relevance and diversity in participants’ backgrounds and experiences as well as the qualitative research design, which entails engagement with participants to elicit rich qualitative insights (Creswell and Poth, 2016). Another criterion for determining the number of participants is theoretical saturation. Theoretical saturation occurs in data collection when new information ceases to emerge and themes become repetitive (Strauss and Corbin, 1998). The key thing is reaching the point of data saturation, where new participants are not providing substantially new insights. Data collection continued even after initial similarities in responses and to balance depth of insight with practical considerations, we stopped reaching out to new participants after we completed 20 interview sessions with prototype walkthroughs. This number has been proven to be sufficient to obtain meaningful results through user studies (Faulkner, 2003). As themes became repetitive, 20 participants provided sufficient diversity and depth, and this was deemed appropriate since

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

the aim of the study is to gain a deeper understanding of participants' perspectives within a specific use case and context (Marshall et al., 2013). The recruitment criteria included 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 sessions were conducted both online via a video conferencing and screen-sharing application as well as in person, depending on the availability of the participants. Before the study commenced, participants received information regarding the aims of the particular study and the overarching goals of the research project, and were also briefed on the voluntary nature of their participation. Additionally, they were provided with an overview of the study's procedures and were reminded of their right to withdraw from the study at any point.

This research study was examined and approved by the Norwegian Center for Research Data (Norsk senter for forskningsdata—NSD) under the reference number: #931033. This approval process involved a comprehensive review of the research protocol, the recruitment process, informed consent procedures, and data handling to confirm that the study adheres to ethical guidelines and legal regulations. The approval confirms that the study has undergone a thorough ethical review and met the necessary standards and requirements, ensuring the protection and well-being of our participants throughout the research process.

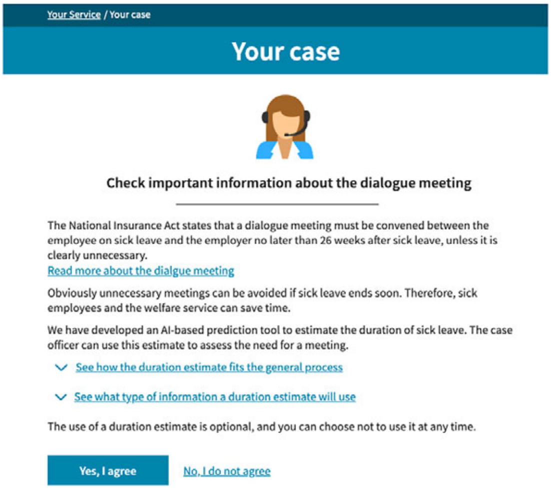
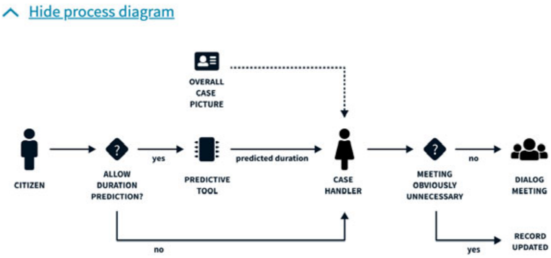
3.3. Data collection

The data collection included three consecutive stages (Table 3). In the initial stage, 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 AI" with an average self-reported rating of 2.20. Overall, the participants had some rudimentary knowledge about AI, but lacked a deeper technical understanding. The second dimension we asked about is: "frequency of technology use." Participants provided an average self-reported rating of 4.65, demonstrating exposure and general familiarity with the use of technologies such as computers or smartphones. Next, we collected data on the level of trust toward the Norwegian government and governmental organizations and on the use of AI within public services. These questions addressed the macro as well as the meso level of the social contract framework. The data collected through these questions provide indications about the current state of the government–citizen relationship and about participants' understanding of the current balance between individual and collective benefits from public services.

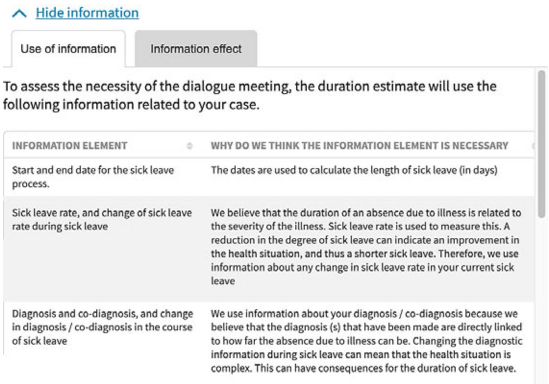
In the second stage, a moderated user study in the form of a task-based interaction with the prototype was conducted. Participants were presented with a short scenario and task that led them to a decision whether they would consent to the use of a new AI-supported prediction system in relation to the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

In the third and final stage, 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 related to the perceived competence and efficiency of the AI-enabled system, the anticipation of negative consequences of its use in public services, and the understandability of the system. Further, we also asked the participants about their levels of comfort and trust toward the specific system they tried. Additionally, we asked open questions about the different explanatory elements used in the prototype.

Table 1. *Prototype overview*

<table>
<tr><th>Prototype</th><th>Description</th></tr>
<tr><td>Figure: Prototype info screen

Your Service / Your case
Your case
Check important information about the dialogue meeting
The National Insurance Act states that a dialogue meeting must be convened between the employee on sick leave and the employer no later than 26 weeks after sick leave, unless it is clearly unnecessary.
Read more about the dialgue meeting
Obviously unnecessary meetings can be avoided if sick leave ends soon. Therefore, sick employees and the welfare service can save time.
We have developed an AI-based prediction tool to estimate the duration of sick leave. The case officer can use this estimate to assess the need for a meeting.
See how the duration estimate fits the general process
See what type of information a duration estimate will use
The use of a duration estimate is optional, and you can choose not to use it at any time.
Yes, I agree
No, I do not agree</td><td>On the info screen, the prototype interface provides textual information about the legal framework for the use case. Further, it presents a value proposition for the use of AI, 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.</td></tr>
<tr><td>Figure: Process chart


Hide process diagram
OVERALL CASE PICTURE
CITIZEN
ALLOW DURATION PREDICTION?
yes
PREDICTIVE TOOL
predicted duration
CASE HANDLER
MEETING OBVIOUSLY UNNECESSARY
no
DIALOG MEETING
no
yes
RECORD UPDATED

</td><td>The first visual 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 of the process (human-in-the-loop).</td></tr>
<tr><td>Figure: Data table

Hide information
Use of information | Information effect
To assess the necessity of the dialogue meeting, the duration estimate will use the following information related to your case.
<table><tr><th>INFORMATION ELEMENT</th><th>WHY DO WE THINK THE INFORMATION ELEMENT IS NECESSARY</th></tr><tr><td>Start and end date for the sick leave process.</td><td>The dates are used to calculate the length of sick leave (in days)</td></tr><tr><td>Sick leave rate, and change of sick leave rate during sick leave</td><td>We believe that the duration of an absence due to illness is related to the severity of the illness. Sick leave rate is used to measure this. A reduction in the degree of sick leave can indicate an improvement in the health situation, and thus a shorter sick leave. Therefore, we use information about any change in sick leave rate in your current sick leave</td></tr><tr><td>Diagnosis and co-diagnosis, and change in diagnosis / co-diagnosis in the course of sick leave</td><td>We use information about your diagnosis / co-diagnosis because we believe that the diagnosis (s) that have been made are directly linked to how far the absence due to illness can be. Changing the diagnostic information during sick leave can mean that the health situation is complex. This can have consequences for the duration of sick leave.</td></tr></table></td><td>The second information element is a table providing an overview of data used by the AI system and an explanation on why different data are needed.</td></tr>
</table>

(Continued)

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

Table 1. *Continued*

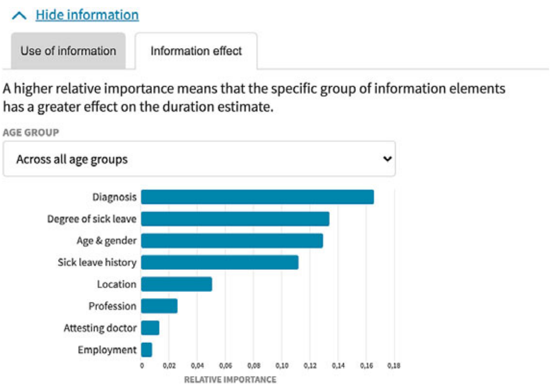
Prototype	Description
Figure: Feature importance chart  Hide information Use of information Information effect A higher relative importance means that the specific group of information elements has a greater effect on the duration estimate. AGE GROUP: Across all age groups Diagnosis Degree of sick leave Age & gender Sick leave history Location Profession Attesting doctor Employment 0 0,02 0,04 0,06 0,08 0,10 0,12 0,14 0,16 0,18 RELATIVE IMPORTANCE	The final information element is an interactive chart, depicting relative feature importance, which aims to provide transparency to the model logic.

Table 2. *Overview of participants*

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

Table 3. *Data collection stages*

Stage	Data collected
1	- Demographic data, i.e., age and gender - Self-assessment for “prior knowledge about artificial intelligence” and “frequency of technology use” - Rating of trust toward “the Norwegian government,” “governmental organizations” and on the use of “AI within public services”
2	- Observational data from interaction with prototype - Think-aloud protocol
3	- Rating of agreement to predefined statements about “perceived competence and efficiency,” “anticipation of negative consequences,” and “understandability of the system” - Open questions about explanatory elements

Table 4. Data analyses classes

Class	Analyses
Scale ratings	- Average ratings for levels of agreement toward predefined statements - Comparison of levels of comfort about AI systems in public service before and after engagement with the prototype
Spoken word	- Theme elicitation based on answers to open-ended questions as well as expressed feedback and comments from the think-aloud protocol during prototype interaction
Observed interactions	- Examination of user behavior during prototype interaction

3.4. Data analysis

To ensure a comprehensive analysis, we categorized the collected data into three classes: scale ratings, spoken word feedback, and interactions with the prototype (Table 4). Each class of data was first analyzed individually followed by a synthetization with the other two classes. The goal of the analysis was to identify themes that would help us to better understand the stances and concerns of participants. Our comprehensive data analysis was performed through the social contract lenses framework. This encompassed an evaluation of power dynamics, a nuanced assessment of collective versus individual interests, and an exploration of how participants perceived and experienced the various agreement mechanisms. First, we evaluated the answers to the predefined questions given by the participants. We began by aggregating the scale ratings in a spreadsheet, which allowed us to create a comprehensive overview of the data's diversity and trends. This also allied us to facilitate the identification of distinct groups, enabled us to gauge participants' perceptions and opinions, and helped us in quantifying and monitoring changes occurring throughout the study.

In a next step, we conducted an in-depth analysis of the qualitative data obtained from both the think-aloud protocol, consisting of spoken responses, and the feedback participants provided in response to open-ended questions regarding various types of explanations. To carry out this analysis, we reviewed the study recordings and thoroughly examined interview transcripts multiple times. The objective was to identify common patterns, perceptions, and recurring themes within the data collected from participants. Throughout this analytical process, we selectively highlighted noteworthy incidents, quotes, and specific expressions shared by the participants and engaged in thorough discussions among the research group members to collectively interpret these findings. This category of data yielded a wealth of valuable information concerning the concerns and understandings of the study's participants.

In a final step, our investigation extended to the examination of participant interactions with the prototype, based on the recordings as well as notes taken during the research sessions. We focused on determining whether specific information elements were accessed, how participants engaged with this information, and the duration of time spent processing it. By merging this analysis with the findings from the previous stages, we aimed to fortify our analytical examination and uncover emerging concepts and recurrent themes through cross-referencing the observations and findings.

4. Findings

The data were analyzed through the social contract lenses framework (Figure 2). Through the macro-lens, we explored the relationship between government (specifically, a governmental organization as the government's executive institution) and the citizens. The meso-lens guided us to examine the concepts of the collective good versus individual good, and the emerging tensions when balancing between them. Finally, the micro-lens oriented our attention to analyzing different types of agreement modeling mechanisms. In the following sections, we present the study findings organized according to these three different lenses defined (Table 5).

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

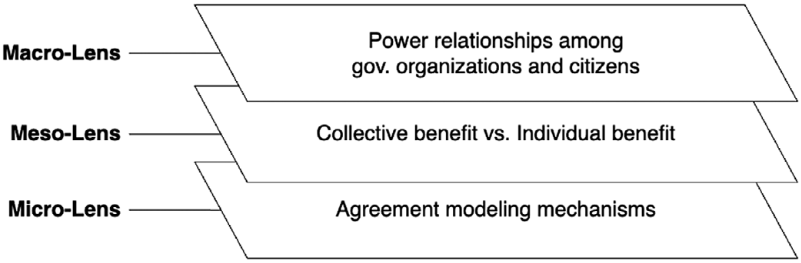

Macro-Lens
Power relationships among gov. organizations and citizens
Meso-Lens
Collective benefit vs. Individual benefit
Micro-Lens
Agreement modeling mechanisms

Figure 2. *Social contract lenses into socially responsible artificial intelligence (AI).*

Table 5. *Findings overview*

Social contract lenses	Key insights
Macro level	Generally positive stance toward AI systems within regional context - High trust in government within regional context - Exposure and engagement with the AI system enhanced the positive stance
Meso level	Mixed perceptions and considerations - Awareness of collective needs and governmental duties - Still concerned about individual benefit
Microlevel	Humans in the loop - Ensures discretion in the decision-making process - Human point of contact available if needed Transparency - Generally appreciated to be open about data usage and processes - Complete understanding of provided information less important

4.1. Macro-lens—Government—Citizen relationship

The macro-lens turns particular attention to the aspects of power dynamics. The power granted to governments by the legitimation of the public enables rulers to fulfill governmental duties. Believing that a government and its institutions will act in the interest of the citizens is an integral part of a social contract and can be seen as a fundamental requirement for a society to work.

A major finding is that among the 20 participants, only one did not 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 toward the use of AI in the public sector. 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 to three categories: skeptic (below 3), neutral (3), and 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 the participants that started neutral converted to being comfortable, one of the two “skeptics” became “neutral,” and two of the “comfortables” became more comfortable.

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

Finally, among the participants who lowered their rating one "neutral" became "skeptic" while the others remained in the "comfortable" category but reduced their expressed level of comfort. According to the framework, these findings relate to an unscathed perception of power relationships between citizens and the governmental organization, from the citizens' perspective.

Exploring the generally positive attitude toward AI in public services further, we found that the participants linked this stance to their overall trust in the government. Several participants provided revealing articulations about their trust in the 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 and its executing organizations. However, some of the participants also expressed reflective comments on how self-aware they are about their level of trust in 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 that manifests a feeling of comfort but also potentially a concern about ensuring that their trust is not betrayed, that is, the social contract breached.

When applying the macro-lens, the existing trust in a government reveals a well-established working relationship between citizens and governmental organizations. The fact that some participants even agree to the use of personal data and AI, without reading through the full information indicates the credibility of public organizations.

4.2. Meso-lens—Interplay between individual and collective good

Another theme that was found in our study concerns the interplay between individual and collective needs and the respective deliberation from citizens. As part of the social contract, citizens are aware of their role as part of a society, but also defend their own individual interests. The meso-lens allows investigation into the different understandings of the individual and the collective good.

We observed participants being aware of the objective of governmental organizations acting in the interest of "the people," with one participant expressing this understanding as *"I really think they are trying to do the best, even if they don't do the best for everyone, all the time."* Another participant reflects on the responsibility to not perpetuate biases, misuse personal data, or discriminate against marginalized groups: *"Whenever you work with AI, there's just all this data that needs to be collected and dealt with and I fear that it's easy to cut corners. And I think you've seen this with the big tech companies—they make something cool and then it turns out that it's racist [...]. Which then makes me concerned that the same problems are not quite solved yet."* Furthermore, participants were also reflecting on the efficient use of public resources, which can be seen as a benefit for society but also, by extension, for the individual citizens in regard to sensible spending on their tax money. One participant mentioned a personal example: *"For my own sick leave, I was so annoyed about all these dialogue meetings, and I thought it's so expensive for the government and it's so unnecessary. How can they think that I should go back to work when they are paying for my "Stråling" [radiotherapy]? It was so stupid and inefficient, so if this [the AI tool] could help to just not spread out all the money it would have been better."*

However, participants were also expressing reflective thoughts from their individual perspectives. For instance, when asked about potential negative consequences of the use of the tool, one participant said: "*No, [...] as long as I have the opportunity to ask for a meeting myself if I want to have one,"* highlighting that the use of such a decision-support system must not restrict any legal individual rights. Another perspective was formulated by a participant, who raised the concern that by pooling the data of the people into groups, there is a risk of losing the intricacies of the individual case, resulting in oversimplification instead of addressing the distinct need of a particular human being: "*It's easy to put people in these boxes on the "Information effects" [referring to the section in the prototype]. So it was*

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

actually the scale on the different age groups, that made me think that this would have a negative effect."

These instances reveal that on the meso-level of the social contract theory framework, citizens continuously shift between considerations about the collective good and the individual need, depending on the context. A clear articulation of the AI value for both society as a whole and individuals is important.

4.3. Micro-lens—Humans in the loop and transparency

Through the micro-lens, the actual processes and mechanisms to reach an agreement among the involved actors is analyzed in detail. This lens shifts attention to the interplay and exchanges between citizens and the public organization, examining the immediate communication and interaction between the different actors to model an agreement. It allowed us to investigate if and how citizens are informed about the intended use of algorithmic models and their personal data and the mechanisms to consent, dissent, or contest.

In our investigation, the micro-level of the social contract theory framework, relates specifically the role of human actors during the process and the perceived transparency. The participants expressed that they feel reassured by having humans involved in decisions alongside AI systems. Within the public service scenario of the study, the leave duration prediction is used as an additional information source for the responsible case worker. It is not an automated decision-making system. The general concept of having humans 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 part of the decision but with people."* Another participant stated: *"It would be much easier 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 ultimate power of decision. However, two participants did express their opinions 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 fairer if it's based on this [AI system], and not if I'm yelling or crying."* This hints toward a perception that having a human within the process may also be a potential weakness that could lead to less fair or equal treatment of citizens.

Some participants related their reasoning for having an HITL to their need to ensure contestability. Participants expressed the concern that if a decision were 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: *"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 cannot AI."* This involvement of a human point of contact and decision maker relates to the microlevel of the social contract lenses framework, as it provides insights into potential agreement modeling mechanisms. It highlights that there exists an expectation to have a human actor available within these processes, rather than a pure system–citizen relationship.

Further, our analysis also revealed findings relating to the transparency provided by the prototype. Although the overall impression of the explanations provided was positive, and the prototype interaction enhanced the positive stance toward AI for most participants, some of them also provided comments that indicate that AI remained relatively opaque for them. Multiple participants commented on difficulties understanding the text. Some of the comments were 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 perceived as hard to understand: *"I don't understand the word information effect."*

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

Similarly, 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."* But 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 existence and availability of information and explanations contribute toward a positive stance about AI, although citizens do not fully understand the information provided.

5. Discussion

The findings of our study show a generally positive stance toward the use of AI in public services. A deeper analysis of the empirical material through a social contract perspective led to the identification of three factors contributing to this positive stance. On a macro level, we identified the high level of trust in the Norwegian government as a contributing factor. On the meso level, we identified the importance of a clear value proposition for AI and how it can benefit collective and individual needs. Finally, on the microlevel, the reassurance provided by having humans in the loop and the perceived transparency into processes and data usage for AI also contributed to the positive perspective. In the following paragraphs, we discuss these findings elaborating on the social contract lenses framework applied.

Trust in a government and its executive bodies to act in the best interests of its citizens forms a foundational aspect of the social contract and is a vital prerequisite for the proper functioning of a society. This trust empowers leaders to carry out their governmental responsibilities and maintain the legitimacy they derive from the populace. Socially responsible use of AI technology in public services relates to questions of power, concerning, for example, the repurposing of existing personal data to train AI models, the usage of personal data to improve administrative processes and decision-making but also about potential misuse or the discrimination of marginalized groups. In our context of being asked to consent to the use of AI for public services, the trust in government expressed by the participants describes the expectation of responsible use of AI and the processing of personal data. In particular, it relates to two types of transparency. First, the expectation of transparency on which data will be used, what is the rationale of the data usage, and for which benefits and purposes. Second, the articulation of a clear value proposition as well as its implications by weighing the public good versus the individual benefit (Schmager, 2022). 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 expectation is upheld, even if the provided transparency and explanations are not fully understood. The 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, constituting a power abuse from the governmental organization (Xu et al., 2006).

This reveals an interesting relationship and tension between the macro level and the microlevel of the social contract lenses framework. From a macro-lens perspective, the sole existence and availability of information and explanations seem to contribute toward a positive stance toward AI, because the perceived transparency and trusted relationship make citizens accept a vulnerability caused by the lack of understanding. At the same time, this prompts questions concerning the essential level of comprehension necessary for informed consent and the shared responsibility of citizens, both as individuals and as integral parts of a society. To what extent is an understanding of the domain and technology deemed suitable or necessary for citizens to make informed decisions? As shown by Bayer et al. (2021), there is evidence indicating that an elevated level of expertise and comprehension in a domain can potentially erode trust in AI-driven decision-making systems. Hence, it becomes a matter for consideration whether increased AI literacy enhances or hinders citizens "trust." Furthermore, this also touches on the expectation that new AI-enabled processes will include safeguards, for example, 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 toward an implicit expectation from the individual citizen to possess the power to exercise their right as the sovereign to question governmental decisions.

Trust and transparency do not manifest uniformly across all regions and societies, as they vary depending on the regional and societal context (Robinson, 2020; Bach et al., 2022; Wilson, 2022). These values have undergone evolution and display distinct characteristics, ultimately shaping the prevailing stances toward governments and their institutions. This raises the obvious question of how to tailor the design processes and considerations for socially responsible AI in the public sector in societies where trust in governments and institutions is lacking. We suggest that transparency mechanisms and practices serve actually as catalysts for cultivating trust in the competence and authority of the government. In other words, transparency in governmental processes and a clear positioning of societal roles contribute to fostering a trusting relationship between citizens and their government. Our findings support this, as we observed a growing comfort with the use of AI technologies in public services after the use of the prototype. In particular, one of the two participants with an initially skeptical stance transitioned to a neutral standpoint, half of the participants who were initially neutral developed a comfortable stance, and two out of the eight initially comfortable participants became even more at ease after engaging with the prototype. Therefore, it could be argued that transparency and clear communication fosters and cultivates trust.

In the context of public services, AI creates the opportunity for governmental organizations to a sensible introduction of new technologies, creating collective benefits for the whole society. Leveraging such technologies can also benefit the organizations by enabling them to enhance services, increase efficiency, and improve processes. In turn, the consent of the people to allow governmental organizations to use AI can contribute to benefits for individuals and society at large. However, high levels of trust can also entail their own perils. 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. Our findings reveal that on a macro level, the working relationship between citizens and organizations contributes to a positive stance toward the use of AI systems in the public sector, but at the same time, when analyzed by the micro-lens, it raises the question of whether such de facto trust is enough to warrant a fully informed agreement modeling process. 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. Especially in the context of AI systems in the public sector, the risks are significant. The realization of a breach of trust can lead to halting public service AI initiatives, even after their launch (Misuraca and Van Noordt, 2020; van Veenstra et al., 2021).

Interpreting the theme of "HITL," through the micro lens of the social contract framework alludes to the perception that a human actor is frequently considered a more appropriate counterpart in the agreement modeling process than a machine. The level of trust in the overall benevolence of the public service organization seems to be extended toward the public service employee, but less toward the AI system. This also touches upon considerations about the role and responsibilities of public servants in a Weber (1964), and whether or not AI systems are capable of possessing agency or the ability to make value judgments. 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 improved objectivity of the AI system seem to be concerned about the same point, as they are expecting fairer and more equal treatment being the result of an impartial

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

AI system. AI adoption in the public sector entails abiding by the social contract that grants legitimacy to its pursuit to maximize public value for all (Rousseau, 1964). 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.

5.1. Contribution to research

Our findings expand extant research on AI adoption in the public sector (Sun and Medaglia, 2019; Wirtz et al., 2019; Mikalef et al., 2022) by foregrounding the government–citizen relationship that has received limited attention in this body of literature. To drive this research, we propose a social contract lenses framework as a theoretical approach for researching socially responsible AI in the public sector, yielding a novel and interesting research direction on ensuring ethical and socially responsible practice in the use of modern technology. A social contract perspective can help analyzing AI in public services at three different levels: (1) macro, examining the power relationships among governmental organizations and citizens; (2) meso, investigating the gauging between the collective good and individual benefit; and (3) micro, exploring different agreement modeling mechanisms.

The overall positive stance of citizens in this study aligns 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 stance toward 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, providing insights on their reasoning behind their stances. The information provided to citizens fortifies their trust to the government, and having an HITL reassures people that it is possible for them to explain their particular circumstances and even contest algorithmic suggestions if needed.

5.2. 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 the boundary conditions for introducing AI-challenging 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 stance toward AI, we provide practitioners with hands-on pointers when considering a socially responsible design and deployment of AI systems in the public service context. Specifically, the study shows the importance of articulating a clear value proposition that takes into account both the individual and the collective interests and also the provision of clear information about the data uses, the logic of models, and the processes followed, including the role of humans in these processes.

5.3. Limitations and further research

This study is exploratory in nature. It responds to calls for research on citizens' attitudes toward AI (Wirtz et al., 2019; Asatiani et al., 2021; Saura et al., 2022) drawing from rich empirical data collected using a combination of closed and open questions and observations of participants' interactions with a prototype. To ensure that policymakers, legislators, industry, and citizens have the opportunity to understand how cultural values interact with policy discussions about technologies such as AI, broader studies going beyond the societal and geographical borders of a single county are needed. The stances of citizens and their acceptance of AI depend on their cultural identity (Robinson, 2020) and especially their overall trust in government, which is particularly high in Norway (OECD, 2022). As identified in prior research, trust is systemic in nature and is invested in the larger system of public and private actors that are associated with AI (Steedman

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

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 an increasing emphasis on instrumental framings of trust as a resource obscuring the mechanisms through which trust in AI might be built (Krüger and Wilson, 2022). Furthermore, a more complete picture can be developed by exploring both citizens' and public servants' stances regarding digital discretion in AI-supported public services. 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). Public servants' perspectives on this need to be also explored. Taking a social contract perspective can expand prior research on digital discretion (Busch, 2019) by shifting attention from the logic of public servants to the dynamics between citizens and public servants.

It is important to point out that taking a social contract perspective opens up areas for further investigating power relations between citizens and the government. In general, the term social contract encompasses a large body of theories and should rather be understood as an approach rather than a specific theory. Some of these theories have come under criticism, for example, for focusing too much on the negotiations of social contracts and less on what such a contract leaves out (Dworkin, 1973). Furthermore, feminists and race-conscious philosophers have highlighted how power and inequalities are also enacted without really the agreement of those affected (Held, 1993; Pateman, 2016; Mills, 2019). Future research will allow for a more nuanced examination of described actors, relationships, and processes and how they may impact the use of AI in the public sector. Additionally, the alternative theoretical approaches set aside for this particular research should not be dismissed entirely for future inquiries into socially responsible AI. Stakeholder theory, institutional theory, and structuration theory have significant promise in offering unique perspectives that may yield novel and complementary insights.

6. Conclusion

Given the impact technology has in affecting the trajectory of society, the interrelations between technology, organizations, and public policy are increasingly implicated when deciding about the deployment of emerging technologies (Bodrožić and Adler, 2022). AI can benefit public organizations by enabling them to enhance services, increase efficiency, and improve processes, which in turn benefits individuals and society at large. But this also highlights the responsibilities of all parties. Our study provides insights into citizens' stances toward the use of AI in public services in Norway. We found a generally positive attitude toward AI and identified three factors contributing to this positive attitude. These factors are (a) the high level of trust in government, (b) the balanced value proposition between individual and collective needs, and (c) the reassurance provided by having humans in the loop and providing transparency into processes, data, and model's logic. By interpreting the findings through the different lenses of a social contract theory framework, we can provide an explanation of these factors' significance. Inherent trust in 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 toward AI, we advance the discourse on the responsible adoption of AI in the public sector.

Acknowledgments. The authors are grateful to the study participants who generously contributed their time sharing their views and insights.

Funding statement. This research is funded by the Norwegian Research Council through the program IKTPLUSS-IKT og digital innovation, project AI4Users, project number 311680. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

Competing interest. The authors declare no competing interests exist.

Author contribution. Conceptualization: S.S., C.G., E.P., I.P., P.V.; Methodology: S.S., P.V.; Data collection: S.S., C.G.; Analysis: S.S., P.V. Writing original draft: S.S.; C.H.; P.V. Writing—review & editing: S.S., C.G., E.P., I.P., P.V.

Data availability statement. The data that support the findings of this study are available from the corresponding author subject to the restrictions of the NSD approval.

References

Allhutter D, **Cech F**, **Fischer F**, **Grill G and Mager A** (2020) Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data 3*, 5.

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 PR**, **Penttinen E**, **Rinta-Kahila T and 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 (JAIS) 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.

Bach TA, **Khan A**, **Hallock H**, **Beltrão G and Sousa S** (2022) A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction 40*, 1251–1266.

Bayer S, **Gimpel H and Markgraf M** (2021) The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems 32*, 110–138.

Bekker S (2021) Fundamental rights in digital welfare states: The case of SyRI in the Netherlands. In Spijkers O, Werner WG and Wessel RA (eds.), *Netherlands Yearbook of International Law 2019: Yearbooks in International Law: History, Function and Future*. The Hague: T.M.C. Asser Press, pp. 289–307.

Benedikt L, **Joshi C**, **Nolan L**, **Henstra-Hill R**, **Shaw L and Hook S** (2020) Human-in-the-loop AI in government: A case study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. New York: Association for Computing Machinery pp. 488–497.

Bérubé M, **Giannelia T and Vial G** (2021) Barriers to the implementation of AI in organizations: Findings from a Delphi study. In *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Bodrožić Z and Adler S (2022) Alternative futures for the digital transformation: A macro-level Schumpeterian perspective. *Organization Science 33*(1), 105–125.

Buchanan JM (1975) *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Busch PA (2019) *Digital discretion acceptance and impact in street-level bureaucracy* [PhD Doctoral Dissertation]. Agder: Universitetet i Agder.

Center for Humane Technology (2023) The A.I. Dilemma. Retrieved 2 May 2023, from https://www.youtube.com/watch?v=xoVJKj8lcNQ&ab_channel=CenterforHumaneTechnology.

Cohen IJ (*1989*) *Structuration Theory: Anthony Giddens and the Constitution of Social Life*, Vol. 1989. New York: St Martin's Press.

Creswell JW and Poth CN (2016) *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. Thousand Oaks, CA: Sage Publications.

D'agostino F (1996) *Free Public Reason: Making It Up as We Go*. Oxford: Oxford University Press on Demand.

de Sousa WG, **de Melo ERP**, **Bermejo PHDS**, **Farias RAS and Gomes AO** (2019) How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly 36*(4), 101392.

Dignum V (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer.

Donaldson T and Preston LE (1995) The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review 20*(1), 65–91.

Doran D, **Schulz S and Besold TR** (2017) What does explainable AI really mean? A new conceptualization of perspectives. Preprint, arXiv:1710.00794.

Dworkin R (1973) The original position. *University of Chicago Law Review 40*(3), 500–533.

Ericsson KA and Simon HA (1984) *Protocol Analysis: Verbal Reports as Data*. Cambridge: The MIT Press.

EU (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Available at https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Faulkner L (2003) Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers 35*, 379–383.

https://doi.org/10.1017/dap.2024.13 Published online by Cambridge University Press

Flak LS and Rose J (2005) Stakeholder governance: Adapting stakeholder theory to e-government. *Communications of the Association for Information Systems 16*(1), 31.

Freeman RE (1984) *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Fu Z, **Xian Y**, **Zhu Y**, **Xu S**, **Li Z**, **De Melo G and Zhang Y** (2021) Hoops: Human-in-the-loop graph reasoning for conversational recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: Association for Computing Machinery, 2415–2421.

Future of Life Institute (2023) Pause giant AI experiments: An open letter. Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (accessed 2 May 2023).

Gauthier D (1986) *Morals by Agreement*. New York: Oxford University Press.

Giddens A (2014) Structuration theory: Past, present and future. In *Giddens' Theory of Structuration*. London, UK: Routledge, pp. 201–221.

Harsanyi JC (1976) *Essays on Ethics, Social Behaviour, and Scientific Explanation*, Vol. *12*. Dordrecht: Springer Science & Business Media.

Held V (1993) *Feminist Morality: Transforming Culture, Society, and Politics*. Chicago: University of Chicago Press.

Herrmann T (2022) Promoting Human Competences by Appropriate Modes of Interaction for Human-Centered-AI. In *International Conference on Human-Computer Interaction*. Cham: Springer International Publishing, 35–50

Heywood A (2015) *Political Theory-an Introduction*, Vol. *4*. London: Palgrave.

Hobbes T (2016) *Thomas Hobbes: Leviathan (Longman Library of Primary Sources in Philosophy)*. Routledge.

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.

Hume D (1963) Of the original contract. In *Essays Moral, Political, and Literary*, Vol. *1741*. Oxford: Oxford University Press, pp. 452–473.

Jørgensen TB and Bozeman B (2007) Public values: An inventory. *Administration & society 39*(3), 354–381.

Jos PH (2006) Social contract theory: Implications for professional ethics. *American Review of Public Administration 36*(2), 139–155.

Joseph RC and Johnson NA (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*. London, UK: Routledge.

Krüger S and Wilson C (2022) The problem with trust: on the discursive commodification of trust in AI. *AI & Society 38*, 1753–1761.

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.

Locke J (2016) *Second Treatise of Government and a Letter Concerning Toleration*. Oxford: Oxford University Press.

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.

Macher G, **Akarmazyan S**, **Armengaud E**, **Bacciu D**, **Calandra C**, **Danzinger H**, **Dazzi P**, **Davalas C**, **De Gennaro MC and Dimitriou A** (2021) Dependable integration concepts for human-centric AI-based systems. In *Computer Safety, Reliability, and Security. SAFECOMP 2021 Workshops: DECSoS, MAPSOD, DepDevOps, USDAI, and WAISE, York, UK, September 7, 2021, Proceedings 40*. Springer International Publishing, 11–23

Marshall B, **Cardon P**, **Poddar A and Fontenot R** (2013) Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems 54*(1), 11–22.

Mehr H, **Ash H and Fellow D** (2017) *Artificial Intelligence for Citizen Services and Government.* Ash Center for Democratic Governance and Innovation at the Harvard Kennedy School, August, 1–12.

Meyer JW and Rowan B (1977) Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology 83*(2), 340–363.

Mikalef P, **Lemmer K**, **Schaefer C**, **Ylinen M**, **Fjørtoft SO**, **Torvatn HY**, **Gupta M and Niehaves B** (2022) Enabling AI capabilities in government agencies: A study of determinants for European municipalities. *Government Information Quarterly 39* (4), 101596.

Mills CW (2019) *The Racial Contract*. Ithaca: Cornell University Press.

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 Research Reports (JRC120399).

Misuraca G, **van Noordt C and Boukli A** (2020). The use of AI in public services: Results from a preliminary mapping across the EU. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*. Athens, Greece.

Nash Jr JF (1950) The bargaining problem. *Econometrica: Journal of the Econometric Society 18*, 155–162.

OECD (2022) *Drivers of Trust in Public Institutions in Norway.* Available at https:doi.org/10.1787/81b01318-en.

Pateman C (2016) Sexual contract. In *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. London: John Wiley & Sons, 1–3.

Pencheva I, **Esteve M and Mikhaylov SJ** (2020) Big data and AI–A transformational shift for government: So, what next for research? *Public Policy and Administration 35*(1), 24–44.

Perreault L (2015) Big data and privacy: Control and awareness aspects. In *Proceedings of the International Conference on Information Resources Management (CONF-IRM)*. Ontario, Canada Available at http://aisel.aisnet.org/confirm2015/15.

Rawls J (1996) *Political Liberalism*. New York: Columbia University Press.

Rawls J (2001) *Justice as Fairness: A Restatement*. Cambridge: Belknap Press.

Rinta-Kahila T, **Someh I**, **Gillespie N**, **Indulska M and Gregor S** (2022) Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems 31*(3), 313–338.

Robinson SC (2020) Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society 63*, 101421.

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

Rousseau J-J (1964) *The social contract (1762)*. London: Penguin.

Saura JR, **Ribeiro-Soriano D and Palacios-Marqués D** (2022) Assessing behavioral data science privacy issues in government artificial intelligence deployment. *Government Information Quarterly 39*(4), 101679.

Schmager S (2022) From commercial agreements to the social contract: Human-centered AI guidelines for public services. In *Proceedings of the 14th Mediterranean Conference on Information Systems (MCIS), Catanzaro, Italy*. Available at https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1009&context=mcis2022.

Scott WR (2004) Institutional theory. *Encyclopedia of social theory 11*, 408–414.

Sein MK, **Henfridsson O**, **Purao S**, **Rossi M and Lindgren R** (2011) Action design research. *MIS Quarterly 35*, 37–56.

Shneiderman B (2020) Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS) 10*(4), 1–31.

Sentralbyrå S (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.

Strauss A and Corbin J (1998) *Basics of Qualitative Research*. Thousand Oaks, CA: Sage Publications.

Sun TQ 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 39*(3), 101714.

van Veenstra AF, **Grommé F and Djafari S** (2021) The use of public sector data analytics in the Netherlands. *Transforming Government: People, Process and Policy 15*(4), 396–419.

Vassilakopoulou P, **Parmiggiani E**, **Shollo A and Grisot M** (2022) Responsible AI: Concepts, critical perspectives and an information systems research agenda. *Scandinavian Journal of Information Systems 34*(2), 3.

Verhulst SG (2021) Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy 3*, e6.

Waardenburg L, **Sergeeva A and Huysman M** (2018) Hotspots and blind spots: A case of predictive policing in practice. In: Schultze U, Aanestad M, Mahring M, Osterlund C, Riemer K (eds.). *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology: IFIP WG 8.2 Working Conference on the Interaction of Information Systems and the Organization, IS&O 2018. IFIP Advances in Information and Communication Technology*, vol *543*. Cham: Springer. https://doi.org/10.1007/978-3-030-04091-8_8.

Weale A (2020) *Modern Social Contract Theory*. Oxford: Oxford University Press.

Weber M (1964) *The Theory of Social and Economic Organization*. New York: Free Press.

Wieringa M (2023) Hey SyRI, tell me about algorithmic accountability: Lessons from a landmark case. *Data & Policy 5*, e2.

Wilson C (2022) Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly 39*(1), 101652.

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". Available at https://futurezone.at/netzpolitik/der-ams-algorithmus-ist-ein-paradebeispiel-fuer-diskriminierung/400147421.

Wirtz BW, **Weyerer JC and Geyer C** (2019) Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration 42*(7), 596–615.

Wright AP, **Wang ZJ**, **Park H**, **Guo G**, **Sperrle F**, **El-Assady M**, **Endert A**, **Keim D and Chau DH** (2020) A comparative analysis of industry human-AI interaction guidelines. Preprint, arXiv:2010.11761.

Xu H, **Teo H-H and Tan B** (2006) Information privacy in the digital era: An exploratory research framework. In *Proceedings of the Twelfth Americas Conference on Information Systems (AMCIS 2006)*, Acapulco, Mexico.

Xu W (2019) Toward human-centered AI: A perspective from human-computer interaction. *Interactions 26*(4), 42–46.

Zanzotto FM (2019) Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research 64*, 243–252.

Cite this article: Schmager S, Grøder CH, Parmiggiani E, Pappas I and Vassilakopoulou P (2024). Exploring citizens' stances on AI in public services: A social contract perspective. *Data & Policy*, 6: e19. doi:10.1017/dap.2024.13