Drobotowicz, Karolina; Kauppinen, Marjo; Kujala, Sari

Trustworthy AI Services in the Public Sector: What Are Citizens Saying About It?

Published in:
Requirements Engineering

DOI:
10.1007/978-3-030-73128-1_7

Published: 01/01/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Drobotowicz, K., Kauppinen, M., & Kujala, S. (2021). Trustworthy AI Services in the Public Sector: What Are Citizens Saying About It? In F. Dalpiaz, & P. Spoletini (Eds.), *Requirements Engineering: Foundation for Software Quality - 27th International Working Conference, REFSQ 2021, Proceedings* (pp. 99–115). (Lecture Notes in Computer Science; Vol. 12685). Springer. https://doi.org/10.1007/978-3-030-73128-1_7

Trustworthy AI Services in the Public Sector: What Are Citizens Saying about It?

Karolina Drobotowicz(✉), Marjo Kauppinen, and Sari Kujala

Department of Computer Science, Aalto University, Espoo, Finland
{drobotowicz.karolina,marjo.kauppinen,sari.kujala}@aalto.fi

The final authenticated version is available online at
https://doi.org/10.1007/978-3-030-73128-1_7.

Abstract. [**Motivation**] Artificial intelligence (AI) creates many opportunities for public institutions, but the unethical use of AI in public services can reduce citizens' trust. [**Question**] The aim of this study was to identify what kind of requirements citizens have for trustworthy AI services in the public sector. The study included 21 interviews and a design workshop of four public AI services. [**Results**] The main finding was that all the participants wanted public AI services to be transparent. This transparency requirement covers a number of questions that trustworthy AI services must answer, such as about their purposes. The participants also asked about the data used in AI services and from what sources the data were collected. They pointed out that AI must provide easy-to-understand explanations. We also distinguished two other important requirements: controlling personal data usage and involving humans in AI services. [**Contribution**] For practitioners, the paper provides a list of questions that trustworthy public AI services should answer. For the research community, it illuminates the transparency requirement of AI systems from the perspective of citizens.

Keywords: artificial intelligence, trustworthy AI, public sector, transparency, qualitative research

1 Introduction

Recent advances in artificial intelligence (AI) have popularized this area of research after an "AI winter", a period of waning public interest in AI [1]. AI is also gaining the interest of public organizations due to the opportunities it creates [2], such as reducing administrative burdens and taking on more complex tasks to enable public-sector employees to focus more directly on citizens' needs [3]. The European Commission has also imagined that AI could be used to serve citizens 24/7 in faster, more agile, more accessible ways [4].

However, some public AI services have already harmed society. The AI Now Institute [5] has reported that multiple deployed AI systems have led to misleading results or violations of civil rights. For example, in the United Kingdom, thousands of immigrants had their visas cancelled due to an erroneous AI system, and in the United States (US) in 2016, AI dramatically lowered the number of home-care hours for people with disabilities without any explanation or possibility to contest its decisions. In 2019, the AI Now Institute published another report [6] documenting cases of automated decision systems used in US public administration. For citizens who were not expecting the use of AI in these cases, they became an unpleasant surprises and decreased their trust [7].

In light of the rise of AI in society and its potentially harmful effects, multiple private and public institutions have published principles and guidelines for ethical AI [8]. However, existing guidelines for ethical AI systems are mostly the results of discussions with industry and academic experts, rarely including citizens' needs and voices.

The goal of this qualitative study is therefore to investigate what kind of requirements citizens have for trustworthy AI services in the public sector. We present findings from 21 interviews with Finnish residents and a design workshop on four public AI services. The data were collected as part of the "Citizen Trust Through AI Transparency" project [9], the goal of which was to provide ethical guidelines for AI usage in the public sector.

The remainder of this paper is structured as follows. First, we review the existing literature on ethical guidelines for AI systems, with a focus on public-service cases. Next, we present our research method and its outcomes. Finally, we discuss the results and limitations of the study and conclude with suggestions for future research.

2 Related Work

Jobin et al. [8] reviewed 84 ethical AI guidelines proposed by industrial and scientific institutions, ten of which targeted the public sector. They found five principles repeated in over half the guidelines: 1) transparency, which aims to increase system explainability, interpretability, or disclosure; 2) justice and fairness, which are connected to mitigating bias and discrimination and enabling challenge or redress; 3) nonmaleficence, which focuses on system security and safety; 4) responsibility, which is often presented alongside accountability and refers to legal liability and integrity; 5) privacy, which mostly relates to data protection and data use and is presented both as a value and a user right.

Across academic guidelines, we found two that focus on interaction with AI systems. First, Amershi et al. [10] presented a set of human–AI interaction guidelines based on documents from industry, scientific AI-design publications, and tests with design practitioners. They suggested how AI systems should behave and what options they should give users during interactions with them. They also mentioned the importance of making systems' functions, performance, reasons, and biases transparent. Second, Rzepka and Berger [11] studied the literature to understand how system and user characteristics influence interactions with systems, finding that transparency positively influences user behavior.

Among guidelines on ethical AI in the public sector, we found two created by research institutes. The Alan Turing Institute [12] presented an extensive set of guidelines in three parts: 1) support, underwrite, and motivate values for a responsible data ecosystem; 2) fairness, accountability, sustainability, and transparency principles for designing and using services; and 3) a process-based governance framework to operationalize these guidelines. The second document came from the Harvard ASH center [3], and explored AI usage in citizen services, suggesting six strategies for the government: 1) make AI part of a citizen-centric program, 2) solicit citizen input, 3) build on existing resources, 4) be data-prepared and tread carefully with privacy, 5) mitigate ethical risks and avoid AI decision making, and 6) focus on augmenting employees, not replacing them.

3 Research Methods

3.1 Overview of the Qualitative Study

We carried out a qualitative exploratory study to answer our research question. We decided to triangulate our data collection because, as suggested by Carter et al. [13], it animated a deeper understanding of the topic and uncovered more detailed answers to our research question. We chose in-depth interviews and a workshop, which are complementary methods according to Kaplowitz and Hoehn [14]. Indeed, during our interviews, the participants felt safer and more focused to share more details, and interactions during the workshop stimulated the participants to share thoughts and needs that did not occur to them during the interviews. The two methods also induced different responses: in the interviews, the participants were more reactive, and in the workshop, they were more creative. Moreover, similarly to the findings of Schlosser et al. [15], the workshop helped uncover broader perspectives on research questions and start topics that are difficult to cover in interviews.

3.2 Study Participants

A total of 21 participants were interviewed (11 women and 10 men). The ages of the participants varied between 18 and 67, with an average of 35. Of the participants, 12 had university degrees, 12 were Finnish, and 9 were immigrants who had stayed in Finland for 3–20 years, with an average of 9 years. When asked for self-estimations of AI knowledge and interest, six of them admitted a poor understanding of AI, eight had a medium level of AI knowledge, three were actively interested in AI, and four were working in the AI field.

Later, eight people participated in the workshop (four women and four men). Six had participated previously in the interviews. Their ages were between 22 and 38, with an average of 28. All had at least bachelor's degrees. Three were born in Finland, and the other five had been in Finland for an average of 6.5 years. One had poor knowledge of AI, three had medium awareness of AI, two were actively interested in AI, and two were working in the field of AI.

3.3 Data Collection

In both the interviews and the design workshop, we used fictional public-service AI examples (Table 1) to help participants understand the scope of possible AI usage and focus their conversations. These sample AI services were generated based on discussions with public-sector representatives in Finland to ensure that they were realistic. The data-collection materials, such as the interview questions and sample AI services, can be seen in the appendix to the master's thesis of the first author [16].

Table 1. Cases used in the data collection.

AI service	Example case	Id
Decision making	AI service that makes a decision about whether the applicant will receive housing	C1
Health prediction	AI service that predicts the mental health problems of citizens and informs a family about it	C2
Impact assessment	Automatic assessment of education impact on pupils, where data collected from children are processed by AI	C3
Fraud detection	AI service used in the social insurance organization to detect financial fraud	C4

Interviews. We conducted 21 interviews between June and July 2019. Each lasted between 45 and 60 minutes. We designed the interview to be semi-structured, because this provides a good balance between deeply understanding the novel topic and avoiding excessive time consumption, as suggested by DiCicco-Bloom et al. [17]. To assess its validity, the interview process was checked by experienced researchers and piloted with an external person whose answers are not included in the results.

The interviews were divided into three parts. First, the participants were asked general questions about their current attitudes and knowledge about AI in the public sector. Second, they were presented 2–4 public-service AI cases from Table 1. Cases were given in a counterbalanced order to avoid sequence bias. The interviewees were asked to share their concerns, needs, and questions related to each case. All the cases were deliberately information-scarce to nudge the participants to point out what they were missing. The third part of each interview contained a few follow-up questions on using AI in public services.

The participants were invited via physical social security offices and online university channels. We aimed to find people with diverse educations, ages, genders, and AI knowledge. The participants had to be over 18 years old and to have lived in Finland for at least three years. If the person was comfortable speaking English, the interview was conducted in that language; otherwise, it was done in Finnish. Each interview was audio-recorded and transcribed with the consent of the interviewees. A movie ticket was given in exchange for participation.

Design Workshop. The workshop was conducted in July 2019. Its goal was to engage citizens in determining requirements for trustworthy public AI services. The workshop method was inspired by the focus group technique [18] and the ideation methodology described by Michanek and Breiler [19]. The workshop started with a warming-up game. Later, the main task was introduced and repeated three times. Each group found an information sheet by their workstation with a description and example case of AI service in the public sector (Table 1). In the groups, the participants were asked to discuss how to make these services trustworthy. They were asked to save the results of their discussions in whatever form they found useful, using any of blank A3 paper, sticky notes, pens, printed phone mock-ups, and markers. The results of each group's discussion were then visible to the next group coming to the workstation so that the participants could be inspired by previous outcomes. Each round, the participants were put in different groups of two or three people at a new workstation. The workshop ended with a follow-up discussion in which the participants summarized all the results. The workshop was two hours long, and each participant was offered two movie tickets for participating.

3.4 Data Analysis

All the collected data were open and axially coded according to *Research Methods in Human-Computer Interaction* [20]. First, we analyzed the interview and workshop data separately. Next, we compared the results of these two analyses and then combined them. Whenever a result would come solely from AI experts, we recorded it and specified in the text below.

Interviews. Transcribed interviews were analyzed with the support of the qualitative data-analysis software Atlas.ti. The analysis started with one researcher reading the transcripts and marking segments of texts with descriptive in vivo codes. Three types of codes were identified: needs, concerns, and questions. We then iteratively categorized and compared coded segments of data. First, we grouped codes of all forms into high-level concept categories, such as "data" and "purpose." For example, the concept "data" included codes such as "datasource" or "consent." Second, for each category, we read all the included text segments and clustered them into subcategories. For example, in "data," we distinguished subcategories such as "data collection" and "data bias."

Design Workshop. Data from the workshop were saved in the forms of physical sticky notes and an audio recording of the follow-up discussion. The analysis of the workshop materials was similar to the interview analysis but did not employ any digital tools. First, we reviewed sticky notes while listening to the audio recording to clarify and add missing information. Next, we clustered sticky notes by the high-level concepts to which they were related. We then examined the notes inside each cluster and divided them into subcategories. Like the interview analysis, this was also done iteratively.

Axial Coding. We compared the subcategories of the interview and workshop analyses and identified similarities, differences, and relationships between them. As suggested by Charmaz [21], this iterative process enabled a deeper understanding of the concepts and thus improved the accuracy of our results.

4 Results

In this section, we present the requirements that our participants shared for trustworthy public AI services. A significant part of the discussions with the participants was focused on transparency, so we start by introducing this requirement. We continue by presenting the participants' detailed questions and requirements grouped into five concepts: purpose, data, core AI process, human involvement, and service overview.

4.1 Transparency

All the participants wanted to know more about the public AI services than was presented in the materials. Regarding motivation, they referred to uncomfortable emotions (e.g., "I fear AI if it collects something that is not said. Transparency throughout the research process is needed, otherwise it can feel bad."), the need to make informed decisions (e.g., "If I'm convinced how they get the results, [...] then I can decide"), and trust (e.g., "They don't have to give me all the information at all times, but [they should] be transparent on how they process the information so that I have more trust in them.").

4.2 Purpose

The participants asked multiple questions related to the purposes of services (Table 2). First, many participants highlighted their need to know the topical purposes of public AI services presented in the study. In the follow-up discussion, one interviewee explained, "So, there should be transparency about purpose. What is the intended purpose? What is the base reason this service exists?" Knowledge about the purposes of the services was especially required in the impact-assessment case (C3), where questions like the following emerged: "What are the targets of the project?"

Table 2. Three questions describing purpose-related requirements.

Subcategory	**Questions**
Purpose	*For what* reason was the public service created?
Benefits	*What* are the benefits that the public service brings?
Impact	*What* impact on users or on society can the public service make?

Several participants asked more specifically about the potential benefits of the public services for them and other stakeholders: "What I am expected to benefit from this information?" (C2) and "Would the children benefit from this? Or parents? What is the benefit of the school?" (C3). A few participants also stated that if a service presented a clear benefit to them, they would be more inclined to use it, even if it was not fully transparent. For example, in the health-prediction case (C2), one interviewee said, “A grandmother’s well-being is more important than where the data come from.”

Lastly, a few inquiries were made about the impacts of the public services in the education-related case (C3). Participants asked, “What is the social impact for the participants?”, “If you are part of the experiment, you would like to know what it is for you in the future. How much does it affect your position in society?” Two participants working actively with AI mentioned that public AI services should increase rather than decrease social justice. They mentioned examples of AI methods, such as scoring and grouping, that should not happen in the public sector; for example, “If a child is from a different background and gets results which seem bad, then they might be put in some special group for slow people. But it could just be a misunderstanding of questions or [a] different background. Then, you’re limiting that child’s abilities to do well in the future.”

4.3 Data

Data collection is an essential part of any AI service, and the study participants had multiple questions about it. We categorized these questions into six subcategories (Table 3). Notably, the participants focused their questions on personal data due to the specifics of the presented cases.

Table 3. Eleven questions describing data-related requirements.

Subcategory	Questions
Data source	*What* is the source of data collected in the public AI service?
Data collection	*When* and *how* were the data about the user collected?
	When was the consent given for collecting this data?
Data purpose	*Why* was this specific information needed?
Data storage	*Where* and *for how long* are the data stored?
Data access	*Who* has the access to the data?
Data bias	*Are* the data biased? *Why*? *How* do they impact the results?

First, many interviewees started by asking about the sources of the data in the public AI services. This was especially relevant to the case of automatically prefilled applications for housing (C1) and fraud discovery (C4). For example, two interviewees mentioned, “I’d like to find out where the information came from. It’s irritating when not told here,” and, “Where do they have the data from?” Because there was no information on the data sources during the interviews, participants shared their own guesses and attitudes. In the first case (C1), participants were mostly sure that it came from other public organizations. They shared a positive attitude about that because, in their opinions, it could make the process easier and faster: “I did it all online. And they brought all of the data [from other institutions]. This is really cool because it saves me a lot of time [...]. I knew exactly where they were getting the data from. So, didn’t bother me." However, the fourth case (C4) incited more controversy, as participants guessed that private companies were the data sources. That case provoked more opposing voices, such as, "Maybe they get my income and spending from my bank? I don't think they should do that because that's crossing the border from the public to the private sector." These outnumbered the accepting voices, such as, "It doesn't hurt even if the information is borrowed from somewhere else."

Several participants also wanted to know more details about the data collection. A few interviewees asked about whether and when they gave their consent to share their data: "Where and when do I consent to this? If I didn't consent, then why are they collecting?" The workshop attendants were also interested in how and from what period data were actually collected. According to them, it was also essential to know why specific data are collected for public services. This was especially relevant in the impact-assessment case (C3), in which data were collected from children. One interviewee asked, "What is the justification behind collecting this much information on my and other [people's] children?"

When data are already collected, they must be stored somewhere; the participants with greater AI knowledge were interested in this topic as well. They asked where and for how long they were stored: "How and where are the data stored, and in what kind of format?" The workshop attendants added that they wanted to know what happened to user data after the services were finished and what organizations or people had access to their data: "[I should] know where the information was going." A few interviewees also asked how they could access their collected data.

Lastly, there was some discussion about data bias, that is, how using unrepresentative data can lead to discriminating results. This topic was rarely started by interviewees, possibly due to their lack of knowledge in this area. During the workshop, three of the eight participants who worked with AI or were interested in it knew what AI bias was. Upon discussion, participants suggested the importance of informing users about possible biases in AI systems, why they emerge, and how they can affect results.

Apart from their questions, many participants stated their requirements related to data. First, they highlighted the importance of consent to share data: "[I want to] decide whether or not I consent to some information being collected on me." Next, some suggested that after data are already collected, they should be able to review them. According to interviewees, this would enable users to notice any problems with the data, such as being too old, missing something, or being wrong: "If they collect the data, you should have some sort of report. You could say when something is missing." A few participants also requested full control over their data: "It should be possible to keep track of where the information goes. So, even though it would mean that I won't be favored in certain decisions, I'd still like to control information that is given." "We should be more aware of what our data [are] being used for. And we should be in more control of switching on and switching off what we do and do not share."

Participants also shared concerns about their privacy and the security of their data. They especially opposed too much information being collected about their children and relatives in cases C2 and C3 and their financial status in C4: "[It] wouldn't be okay to see messages they send to the family example," "knowing that your grandma might be in danger of social exclusion just sounds like there's a constant surveillance on her," and "it feels like a privacy violation." They also shared the requirement of storing personal data securely: "Security is really important in a lot of these. Because [...] it can also be exploited by companies to do targeting. Or it can be exploited by the government." "There should be an assurance that no one can access your information." One interviewee with extensive AI knowledge shared another concern: "The only thing that worries me in the public sector is that: Do we have the best people to keep the data protected?".

4.4 Core AI Process

We distinguished the core AI process inside each public service responsible for creating its results, that is, the intentional output generated by each service, such as a decision or prediction. In this section, we describe the subcategories that we grouped around the concept of the core AI process (Table 4).

Table 4. Seven questions describing core AI process related requirements.

Subcategory	Questions
AI process reason	*What* is the reason for using AI in the public service?
Used criteria	*What* criteria are being used for the results creation?
Used data	*What* data are used for results creation?
Results creation process	*What* is the process of results creation?
Results explanation	*What* is the reason for the results? *Which* data and criteria affected the results?
Results reliability	*How* reliable are the results?

First, the participants requested to know the reason for using each AI process, especially in the impact-assessment case (C3): "Why this way? What is the justification behind collecting this much information on my children and other children? And why does it have to be [this way]?" It was also of interest to the participants to know what criteria were used to create the results: "I would certainly be very interested in what the parameters are that affect the decision" and "What kind of laws [do] they have for [the] particular benefit that I'm applying for?" Next, many participants asked what data were actually used to create the results: "What information would be utilized?" and "It would be useful to know what information is used." They also asked much more detailed questions about the data, which are presented above (Section 4.3).

Most of the participants asked about the AI process, that is, what is actually done with the criteria and data to create the results. For example, interviewees asked, "How do they do [the process]? How did they use your data?" "What kind of conclusions are they trying to get out of it?" and "[It would be] good to know how they analyzed this case." According to the participants, it was vital that they at least know that AI is used in the process: "I think there's no reason to hide that [AI is used] because I think some people certainly would have negative feelings if they didn't know" and "I feel awkward; I was tricked. [. . .] If I know that it's an automated process, I will feel better."

The participants also mentioned the need to control the process. During the workshop, participants suggested that each service should have options, such as always being able to quit the service or to have humans handling tasks instead of an AI. Interviewees also shared their worries over not controlling services and AI in general: "I don't think I can control [what happens in the service]" and "I don't think I can stop or make [AI] more humane; it's going too fast."

When participants were given the results of the AI services presented in the study, they often asked for explanations, especially in the decision-making case (C1). Two responses were, "What are the exact reasons?" and "There's a lack of information as to why my application is rejected." A few participants suggested how the explanation should look: "It would need to be professional and have clear indications what the exact reasons are and reference to certain clauses" (C4), "Something like, we have a list of people that have been waiting for a long time, or the refugees, some reason you can understand" (C1), and "Why is it rejected? Like, [...] there is no free apartment. My wishes are too big" (C1). Lastly, one participant commented on the reason for an explanation: "[I would] have to call somebody to try to figure out why. Then they also have to try to figure out why. So, if it is smart enough to decide immediately why I'm not going to get the house, it should also be smart enough to tell me immediately why."

Participants more experienced in AI also requested knowledge about result reliability, that is, their accuracy and trustworthiness. One asked, "How confident are the results? Like, are they 110% confident? Or is it more like it might be that the system works, or it's like 100% prediction, or in a hundred thousand cases before me, this happened?" During the workshop, participants also suggested that especially in the healthcare services it should be clearly written how much the results could be trusted, such as by stating, "This is not a diagnosis and does not replace medical professionals." They affirmed that it is vital to provide levels of confidence in results, as they might be erroneous.

4.5 Human Involvement

Human personnel are usually involved in public-service operations, but they can also be involved in core AI processes. The need to know where actual humans are involved in creating results was highlighted in the workshop, although few interviewees asked for it. However, interviewees shared multiple needs and concerns related to the roles of humans in AI services.

First, the interviewees shared the general need to interact with people instead of AI, especially in personal cases, such as healthcare. One interviewee said, "We can replace as many things as possible with machines. But at the end of the day, we still crave human interaction in some form or another." According to the participants, human personnel could be responsible for introducing people to

Table 5. One question describing human-involvement related requirements.

Subcategory	Questions
Human involvement	*What* is the role of humans in the results creation process?

the service or explaining its results: "If someone is telling face-to-face, it's easier to motivate or convince the person. But if it's some odd papers, sometimes you just skip the part that you didn't need." During the workshop, it was also mentioned that there should be an easy way to contact someone from the service.

Second, the interviewees were concerned that AI would not be able to understand a human case as well as another human, as it overgeneralizes and lacks human intuition: "Especially in healthcare, I want humans to talk to because there are a lot of things that are not possible to be read by the program." In application forms, as in C1, one person suggested such a solution: "I would honestly prefer that there was an open field to describe your life situation right now. And then there would be a human in the loop looking at the application."

A few participants suggested that it would be better to have humans make final decisions, especially when they are important. Such comments were given in the follow-up discussion and on the fraud-detection case: "It is worrying if solely AI would be taking decisions on humans' lives" and "I would like a person to see and decide based on this information rather than artificial intelligence" (C4). One person highly educated in AI added, "As of now, we are still stumbling upon training AI to the point that it does the decisions correctly [...]. I still don't feel comfortable [with an AI] making the decision on its own."

Lastly, several participants suggested having human controllers monitor AI services for possible errors, rule-breaking, and unethical actions. One interviewee who worked with AI said, "Nothing should be [fully] automated. When it comes to analysis and evaluation, you have to have someone who can verify that the system is working according to rules and ethical guidelines, as demanded by society." The workshop participants suggested always having the option to ask for a human review of an AI process.

4.6 Service Overview

The service overview contains general, practical knowledge about each service that participants asked about (Table 6). First, the study participants mentioned their interest in understanding the high-level processes of the public services. Some interviewees asked general questions: "How does this service work in practice?" Others asked more case-specific questions: "How frequently and how will it be available?" During the workshop, participants requested basic information about the service stages and how long they take. They also requested updates on service statuses when results were not immediate, such as in C3. Two interviewees said, "Maybe every half a year, or maybe even once a month" and "It would be good every six months to get follow-up information."

Table 6. Five questions describing high-level service related requirements.

Subcategory	**Questions**
High-level process	*What* should customers expect from the public service? *What* are the stages of the service and *how long* do they take?
Accountability	*Who* is accountable for the public service?
Users of the service	*Who* are the users of the public service?

Some interviewees were also interested in knowing the other users of the service. For example, two interviewees asked about the number of other children whose data would be collected for the educational impact-assessment case (C3): “Is it only my child? Is it the whole class?” and “I would like to know how many other children are involved.” Lastly, participants asked for who or what organization was accountable for the service and its outcomes: “Who has decided?”, “Who has developed it?” and “What is this social welfare organization?”

5 Discussion

5.1 Transparency

The most important finding of this study is that transparency is a critical requirement for trustworthy AI services from the perspective of citizens. This result is consistent with the finding of Jobin et al. [8], who report that transparency is the most common principle across ethical AI guidelines. However, it can be a demanding task to specify the transparency requirement systematically in practice. For example, multiple transparency definitions have been proposed in the AI services context. In this study, we focused on the visibility of the service information and justifiability of AI service processes and outcomes, as defined by Leslie [12] and Turilli et al. [22]. In more detail, Turilli et al. [22] suggested that transparency should explain the processes accomplished by the service (how, by whom, and what was collected and done), as it enables checking whether the service is a product of ethical processes. Hosseini et al. [23] suggested that to reach meaningful transparency, services must be open about policy (why), process (how), and data (what). Our study contributes to these by providing 27 detailed questions that should be answered by AI services in the public sector for citizen trust. We discuss those questions below.

First, the participants were interested in the purposes of the services, why they existed and what impacts they had on them and others. This was especially important when their benefits were not clear. Second, the participants asked multiple questions about data: from what sources and how the data were collected and whether data owners consented to give the data. They also had privacy-related questions, such as who could access their personal data and how they would be stored. Only a few participants raised the topics of bias and fairness, even though it was one of the most common principles found by Jobin et al. [8], perhaps indicating that those topics are not well known among non-specialists.

Third, the participants shared multiple questions about core AI processes. They were interested in what data and criteria were used and how they were processed to create results. This information was relevant for participants both before they joined a service and as an explanation of its results. This supports the findings of Chazette et al. [24], who found that the vast majority of their survey respondents found service-result explanations necessary. Furthermore, they found that "what" and "why" questions were more important in explanations than "how." A few participants in our study also specified that explanations must be easily understandable by non-specialists, a requirement pointed out in the public-sector guidelines from Alan Turing Institute [12]: explanations should be socially meaningful and devoid of technical language. Fourth, the participants asked questions about the roles of humans in creating results, and fifth, they asked about service overviews.

The results of this study show that AI transparency is very closely related to AI explainability, which has been studied extensively. For example, Arrieta et al. [25] performed the literature review of approximately 400 publications related to expainable AI and defined explainability as "the details and reasons a model gives to make its functioning clear or easy to understand." They also presented explainable AI as a core element needed to achieve responsible AI principles, including transparency. Similarly, Chazette et al. [26] discovered that explainability was the means to achieving the non-functional requirement of transparency. Our study revealed the detailed citizen requirements for explainable AI, such as the visibility of the criteria and data used by the AI and the understandable explanation of results produced by AI.

5.2 Other Requirements

Apart from transparency, the participants shared other requirements. We discuss the two most important here. First, they highlighted the need to have humans involved in services, although participants' views on this diverged. Some emphasized being able to interact with a person to discuss a service, while others only wanted people to be involved in reviewing their data and making decisions or in monitoring the whole AI process. Second, the participants required a certain level of control over their data. Most often, they wanted to be asked for consent before any of their data were collected or shared. A few also requested full control over their data, to be able to choose which data are used, and to be able to withdraw them at any point. Part of these requirements are mentioned in the Harvard ASH Center's strategies for government and public institutions [3], which state that asking citizens for consent to use their data in services creates fewer privacy concerns, discourage letting only AI make critical decisions for citizens, and encourage human oversight.

5.3 Study Limitations

Generalizability. We interviewed only residents of the Metropolitan Area of Finland between the ages of 18 and 67. Despite our efforts to include diverse participants, we cannot confirm that other demographic groups would have similar requirements. In fact, in different parts of the globe, societies have different cultural biases and values that influence their mental models of AI [27]. Even within Europe, cultural and social characteristics vary [28]. Finland, for instance, enjoys greater trust in the public sector [29].

Reliability. We are also aware of the inherent weaknesses of the interview and workshop techniques. For one, the interviewers may have passed their occupational biases into the research. Interviewees also may not have told the truth or not understood the questions well. However, we took a few precautions to counter these threats to reliability. First, our interview questions were reviewed by senior researchers and were piloted. Second, the data-analysis process was reviewed by another senior researcher. Third, the participants came voluntarily for the interviews, they did not need to answer every question, and they were informed that what they said would remain confidential and anonymous.

Six participants took part in both the interviews and the workshop, and we are aware of the bias they may have brought to the workshop by changing or emphasizing opinions they stated during their interviews. However, because of diverse interactions during the workshop, these six participants had chances to discuss topics not covered during the interviews. Moreover, we believe this diversity in the topic awareness likely positively affected results of a workshop by inducing more perspectives to the discussions.

Lastly, we included four AI specialists and three people actively interested in AI in our study. To reason it, we need to share the Finnish AI context. In 2018, Finland released the estimation that one-fifth of its population would eventually need to obtain AI skills [30]. By now, more than 1% of Finnish citizens have expanded their knowledge of AI by taking the freely available course "Elements of AI," and Finnish universities altogether offer 250 AI courses, which are taken by about 6,300 students every year [31]. Finally, we believe those AI specialist are citizens, whose voices are also valuable and who may actively shape future AI in the public sector. For clarity, we also marked all the results that came from only this group of participants.

6 Conclusions

This paper presents citizens' requirements for trustworthy AI services in the public sector. Based on our findings, transparency is a particularly important requirement of public AI services. Specifically, for practitioners, this paper provides a list of 27 questions that ought to be answered by such services to achieve trustworthiness. The results of this study also indicate that citizens have other important requirements, such as the need to control one's data and to have humans involved in AI processes. We suggest that these questions and requirements guide public AI service design and development. For the research community, we contribute by extending the knowledge of the transparency requirement of AI systems from the perspective of citizens.

Reflecting on our experience, we suggest the following for future research. The findings of this paper could be tested with citizens in the form of public AI service prototypes to validate our results and study the depth of information required by citizens to optimize transparency. As another direction, the study of citizen requirements could be broadened by including the private sector. For example, the healthcare sector may be an interesting area to study, as it includes both private and public organizations and is proximal to citizens.

Acknowledgements We thank the Saidot team from spring 2019 for starting the project and assisting in the data collection of this study, J. Mattila for co-organizing and conducting parts of the interviews, and our participants for sharing their experiences.

References

1. Fast, E., Horvitz, E.: Long-Term Trends in the Public Perceptionof Artificial Intelligence. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 963—969 (2010).
2. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M. and Floridi, L.: Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science and Engineering Ethics **24**, 505–528 (2017).
3. Mehr, H.: Artificial Intelligence for Citizen Services and Government. Harvard Ash Center Technology & Democracy (2017).
4. AI HLEG, "Policy and investment recommendations for trustworthy AI," European Commission (2019).
5. AI Now Institute: AI Now Report 2018. (2018).
6. AI Now Institute: Automated Decision Systems Examples of Government Use Cases. (2019).
7. New York City's algorithm task force is fracturing, https://www.theverge.com/2019/4/15/18309437/new-york-city-accountability-task-force-law-algorithm-transparency-automation, last accessed 2020/11/6.
8. Jobin, A., Ienca, M., Vayena, Effy: The global landscape of AI ethics guidelines. Nature Machine Intelligence **1**(2), 389–399 (2019).
9. A Consortium of Finnish organisations seeks for a shared way to proactively inform citizens on AI use, https://www.espoo.fi/en-US/A_Consortium_of_Finnish_organisations_se(167195), last accessed 2020/11/6.
10. Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.: Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13 (2019).
11. Rzepka, C., Berger, B.: User Interaction with AI-enabled Systems: A Systematic Review of IS Research. In: Proceedings of the 39th International Conference on Information Systems, 13–16 (2018).
12. Leslie, D.: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Alan Turing Institute (2019).

13. Carter, N., Bryant-Lukosius, D., Dicenso, A., Blythe, J., Neville, A.: The use of triangulation in qualitative research. Oncology Nursing Forum **41**(5), 545–547 (2014).
14. Kaplowitz, M., Hoehn, J.:. Do focus groups and individual interviews reveal the same information for natural resource valuation?. Ecological Economics **36**(2), 237–247 (2001).
15. Schlosser C., Jones S., Maiden N.: Using a Creativity Workshop to Generate Requirements for an Event Database Application. In: Requirements Engineering: Foundation for Software Quality **5025**, (2008).
16. Drobotowicz, K: Guidelines for Designing Trustworthy AI Services in the Public Sector., Master's thesis, Aalto University, Department of Computer Science, `http://urn.fi/URN:NBN:fi:aalto-202008235015`, (2020).
17. DiCicco-Bloom, B., Crabtree, B.: The qualitative research interview. Medical Education **4**(4), 314–321 (2006).
18. Kitzinger, J.: Qualitative Research: Introducing focus groups. British Medical Journal **311**(7000), 299–302 (1995).
19. Michanek, J., Breiler, A.: The Idea Agent: The Handbook on Creative Processes. 2nd edn. Routledge (2013).
20. Lazar, J., Feng, J., Hochheiser, H.: Research Methods in Human-Computer Interaction. 2nd edn. Morgan Kaufmann (2017).
21. Charmaz, K., Hochheiser, H.: Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Thousand Oaks (2006).
22. Turilli, M., Floridi, L.: The ethics of information transparency. Ethics and Information Technology **11**(2), 105–112 (2009).
23. Hosseini M., Shahri A., Phalp K., Ali R.: Foundations for Transparency Requirements Engineering. In: Requirements Engineering: Foundation for Software Quality. (2016).
24. Chazette, L., Karras, O., Schneider, K.: Do End-Users want explanations? Analyzing the role of explainability as an emerging aspect of non-functional requirements. In: Proceedings of the IEEE International Conference on Requirements Engineering, pp. 223–233 (2019).
25. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion **58**, 82–115 (2020).
26. Chazette, L., Schneider, K.: Explainability as a non-functional requirement: challenges and recommendations. Requirements Engineering **25**(4), 493–514 (2020).
27. Schaefer, K., Chen, J., Szalma, J., Hancock, P.: A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. Human Factors **58**(3), 377–400 (2016).
28. Lee, J., See, K.: Trust in Automation: Designing for Appropriate Reliance. Human Factors **46**(1), 50–80 (2004).
29. Leading the way into the age of artificial intelligence Final report of Finland's Artificial Intelligence Programme 2019. Publications of the Ministry of Economic Affairs and Employment (2019).
30. Koski, O: Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics. Publications of the Ministry of Economic Affairs and Employment of Finland (2018).
31. Artificial Intelligence From Finland, e-Book of Business Finland, `https://www.magnetcloud1.eu/b/businessfinland/AI_From_Finland_eBook/` (2020).